Karlsruhe Institute of Technology

Grid Computing

Andreas Heiss
Steinbuch Centre for Computing

Forschungszentrum Karlsruhe
in der Helmholtz-Gemeinschaft

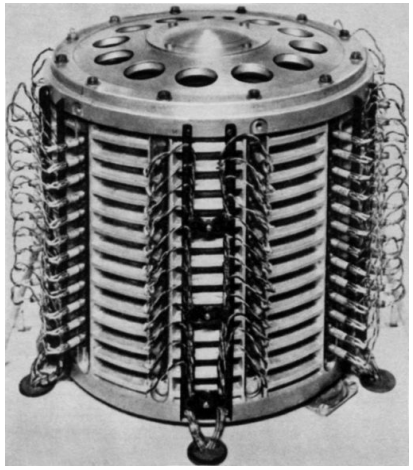
Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825

Computing - the past

Quellen: http://de.wikipedia.org/wiki/Konrad_Zuse

- Zuse 11
  - One of the first commercial computers, available from 1956
  - Price: 120000 DM
    - Taking inflation into account, this corresponds to 600000 € today.
  - 654 relais, floating-point unit
  - 10-20 operations per second

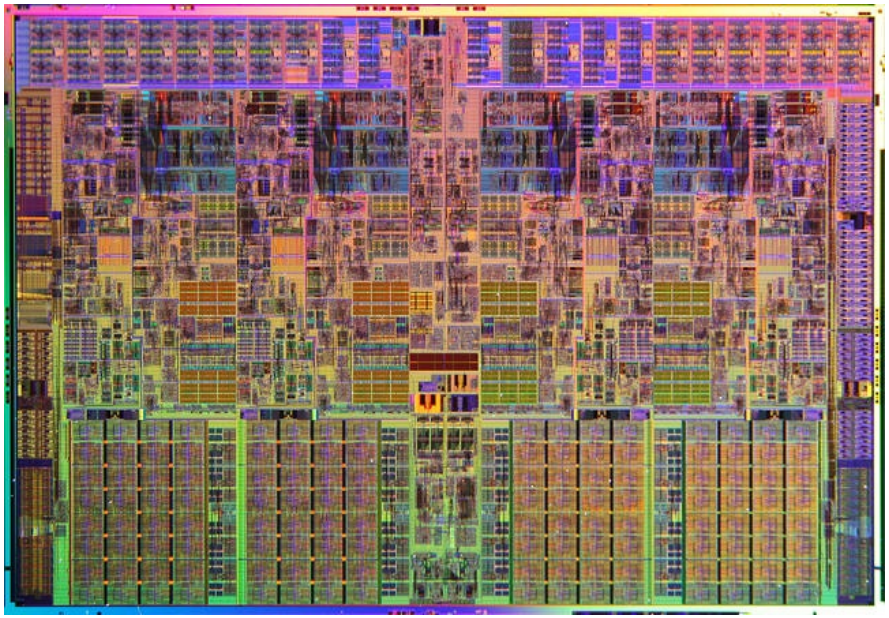
- Drum storage (50s ~ 60s)
  - <10 Mbit capacity
  - 10 Mbit/s transfer rate (fast compared to the capacity!)



Quelle: http://de.wikipedia.org/wiki/Trommelspeicher

- No networks

Computing - today

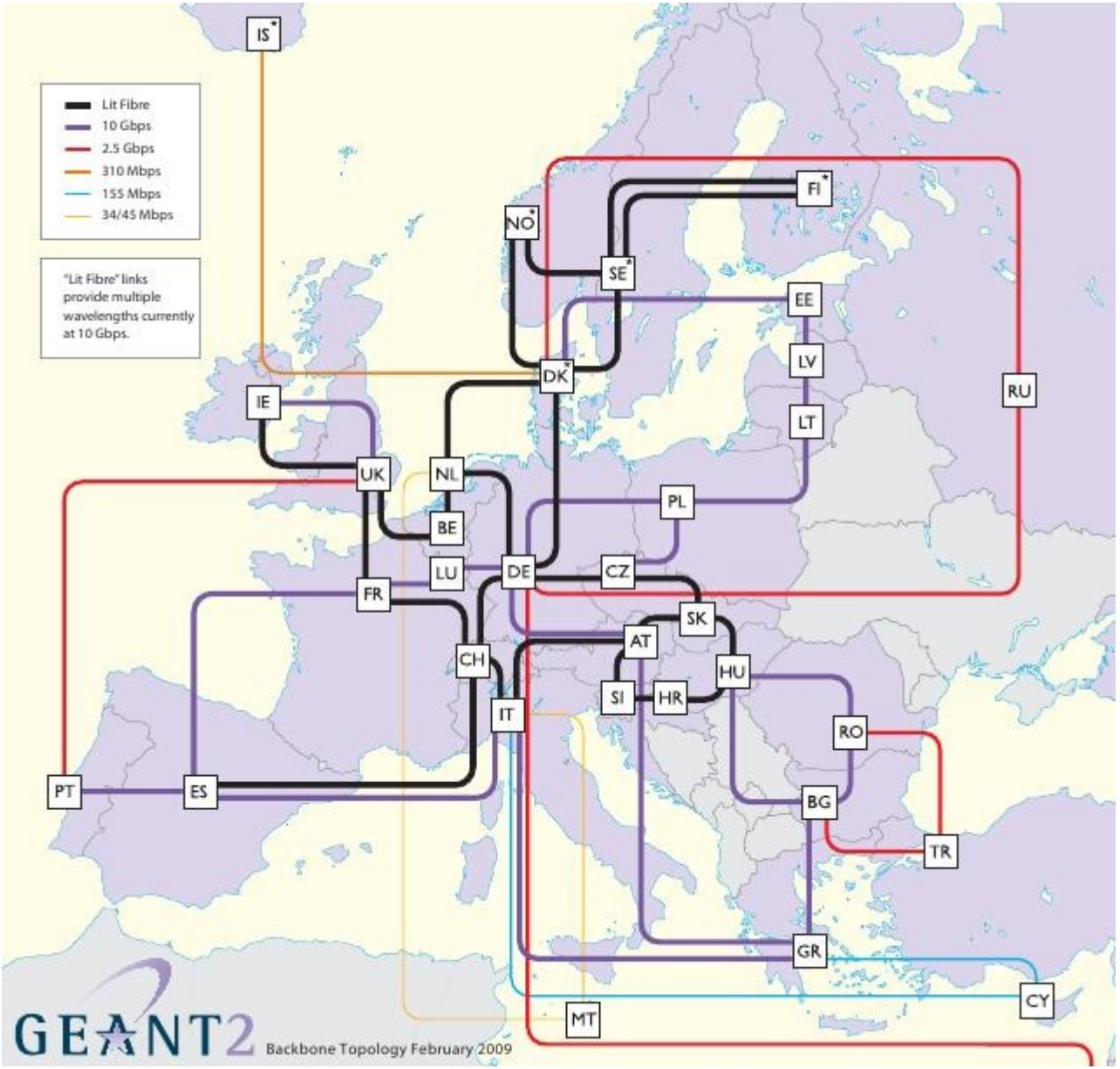


Quelle: http://wwww.intel.com

- Example: Intel Core i7 (”Nehalem”)
  - 45nm structures
  - 4 cores, 8 threads
  - ~ 2.0 - 3.0 GHz
  - ~12 GB/s memory bandwidth
  - 1500 € for a well equipped system

- Online storage: hard disks
  - up to 2 TB per drive
  - seek times < 5ms
  - read/write performance <=100 MB/s
  - ~ 100 € / TB (low performance, desktop)
  - ~ 1000 € / TB (highest speed, 24x7, server)

Computing - today



- **Wide area networking**
  - **Arpanet (Advanced Research Projects Agency Network)**
    - **started 1969, 4 institutes**
    - **~ 50 kbit/s**
  - **Internet today : ~ 10 Gbit/s backbone**

Internal data links:

- **PCI: < 5 Gbit/s**
- **PCI-Xpress: < 64 Gbit/s**
- **QPI: ~ 100 Gbit/s**

Computing - development

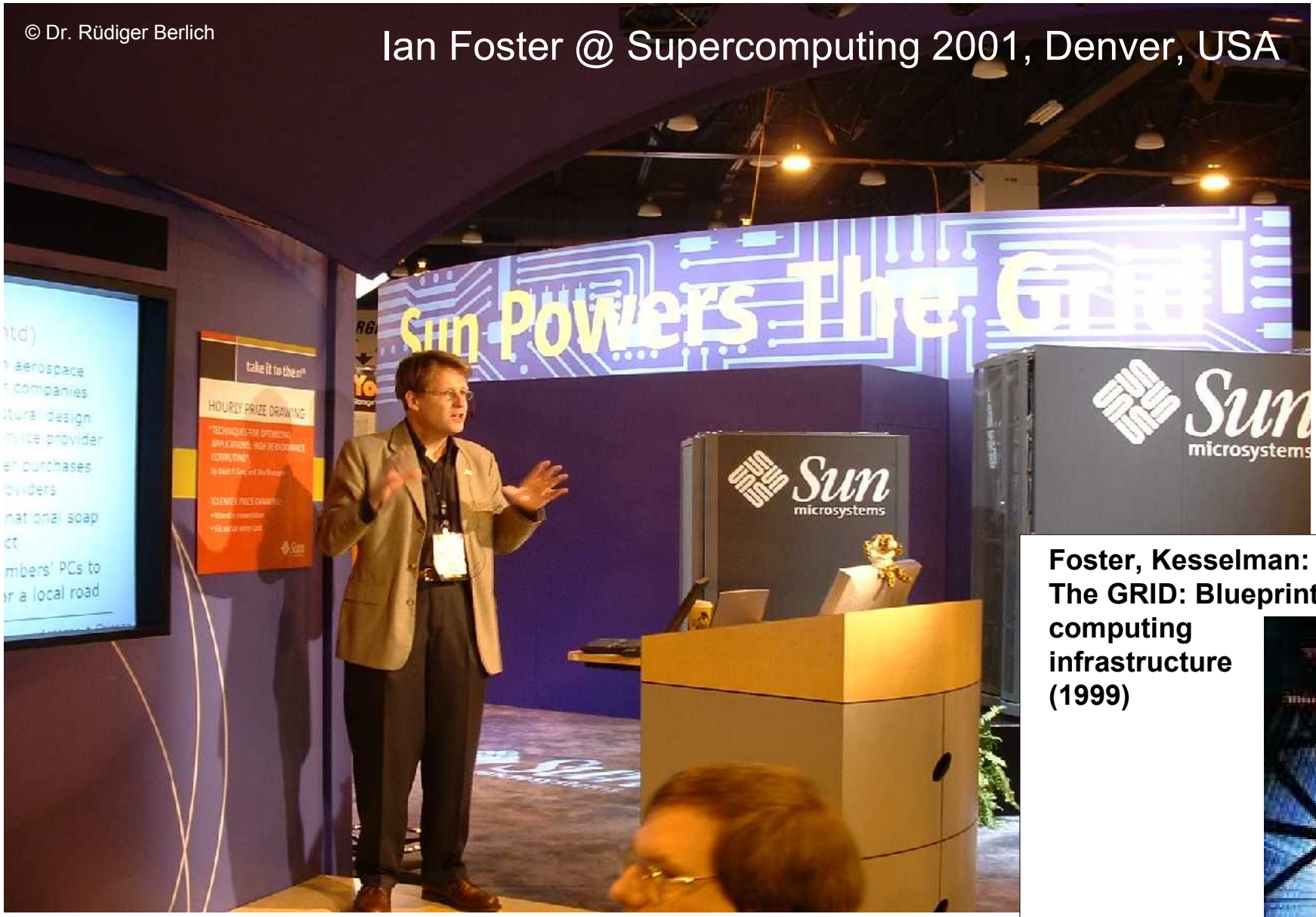
- No. of transistors (and also compute power per chip) doubles every 18-24 months. (Moore's law, 1965/1975)
- Storage densities increase by a factor of 1.5 - 1.8 per year.
- Gilder's law: ”The total bandwidth of communication systems triples every 12 months.”

Computing - development

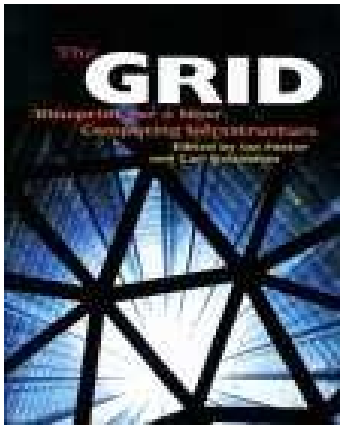
- Available compute and storage resouces are growing almost exponentially.
- Network bandwidths come closer and closer to the speed of internal data links (e.g. PCI).
- External resouces are accessable almost as fast as local resources.

- In such an environment, it is obvious to do certain computing tasks on external resources!
  - No need to buy and operate special devices for each special task.
  - Share resources and improve utilisation
    → minimise costs

Grid Computing



Foster, Kesselman:
The GRID: Blueprint for a new computing infrastructure (1999)



Grid Computing

- *”Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations” (I. Foster)*

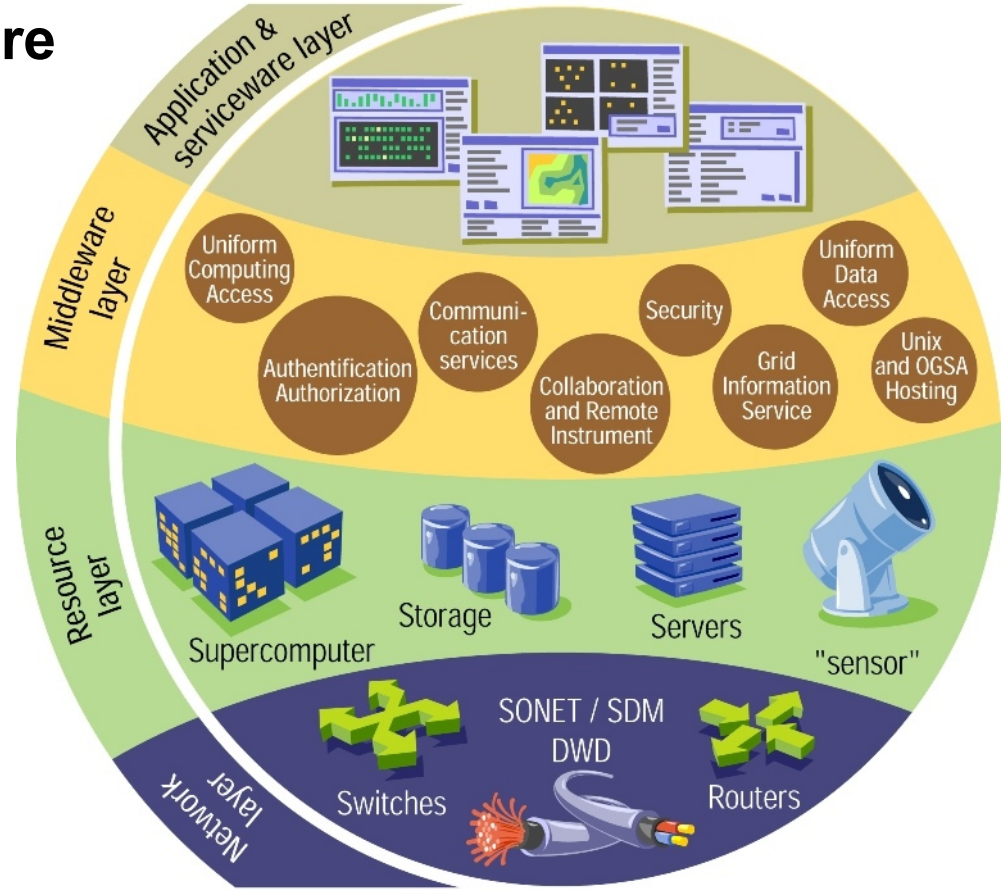
- *”Grid” derives from ”power grid”*

  The vision:

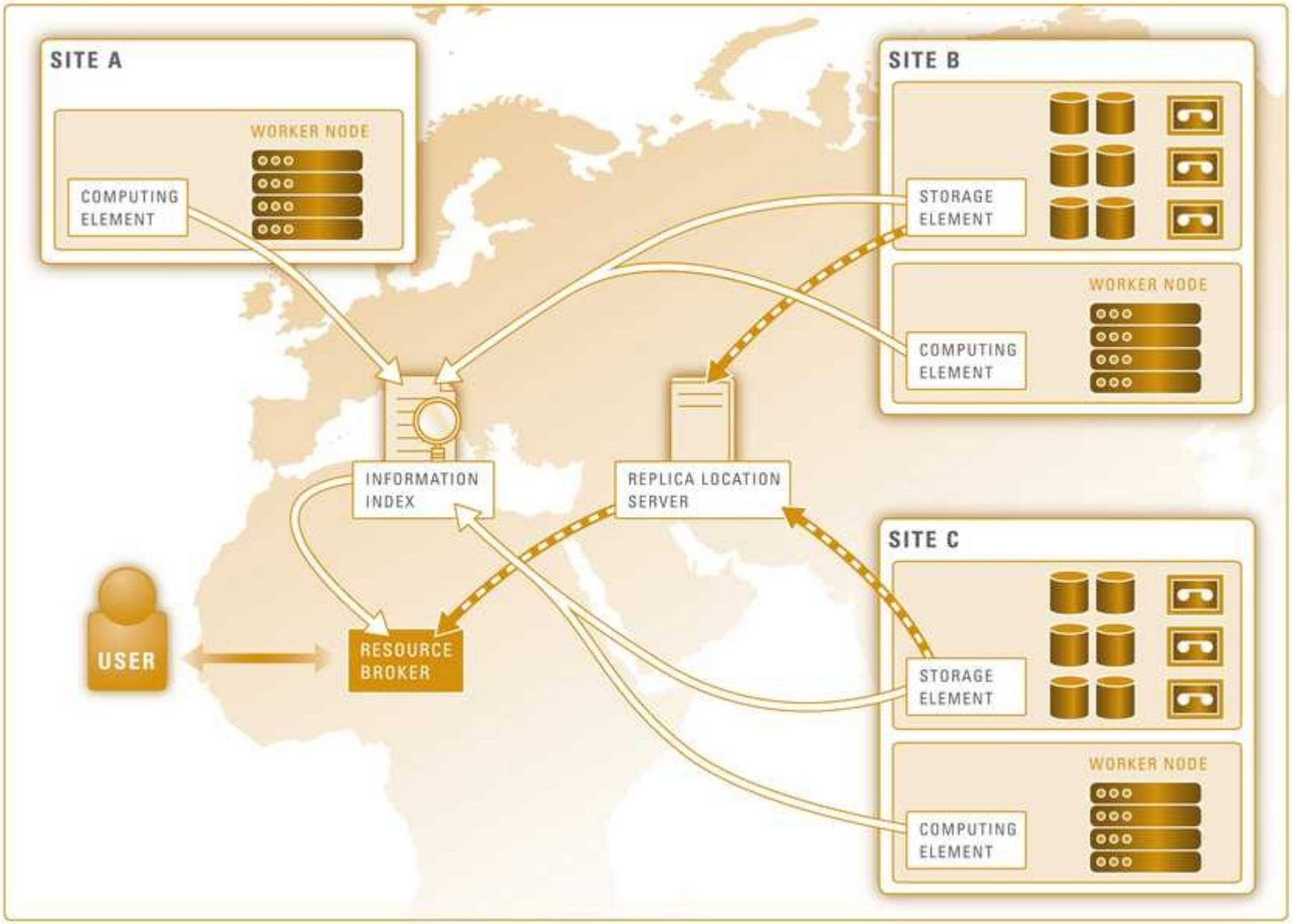
  - *Get computing power from an (computing) power outlet*
  - *Not only compute power, but also storage, access to measuring instruments, sensors, ...*
  - *Just plug in and access resources worldwide*

Grid Computing

- **Middleware** is the 'glue' software that pools together various resources and services and to create the Grid.
  - Interfaces to
    - access compute power
    - access data
  - authentification and authorization (PKI, X.509)
  - information service
  - accounting
  - ...

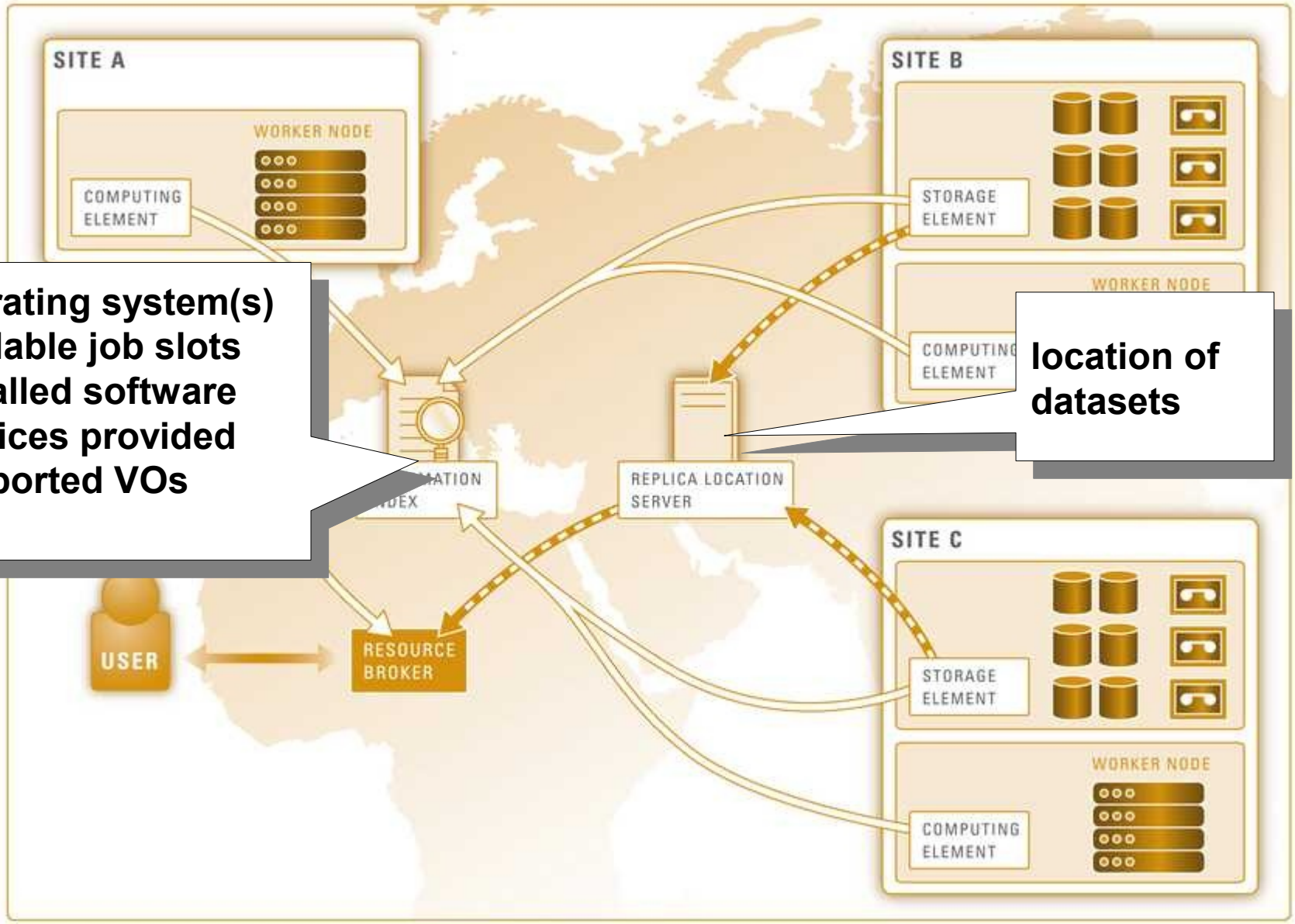


Grid Computing



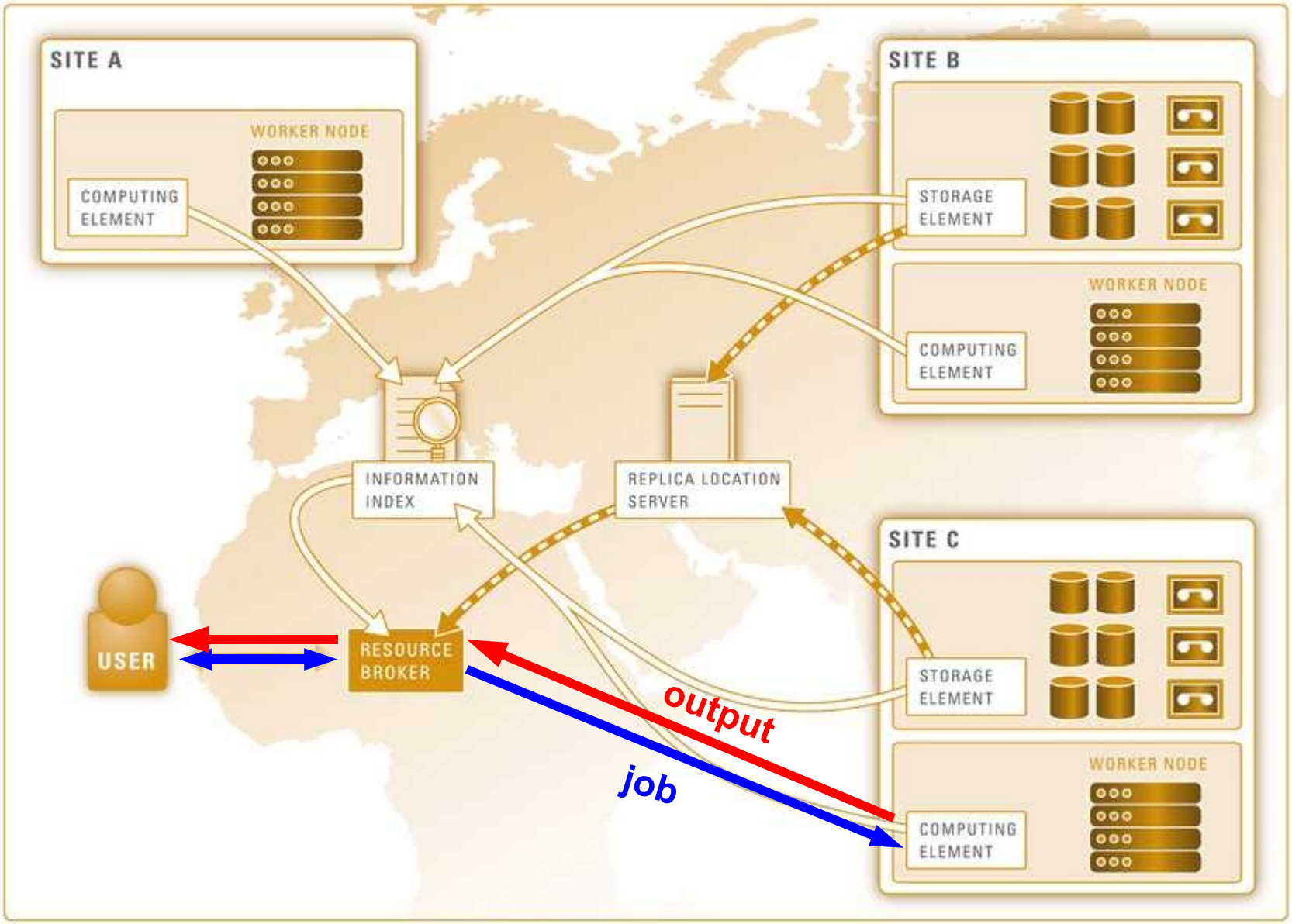
Plot courtesy Martina Hardt / Desginal

Grid Computing



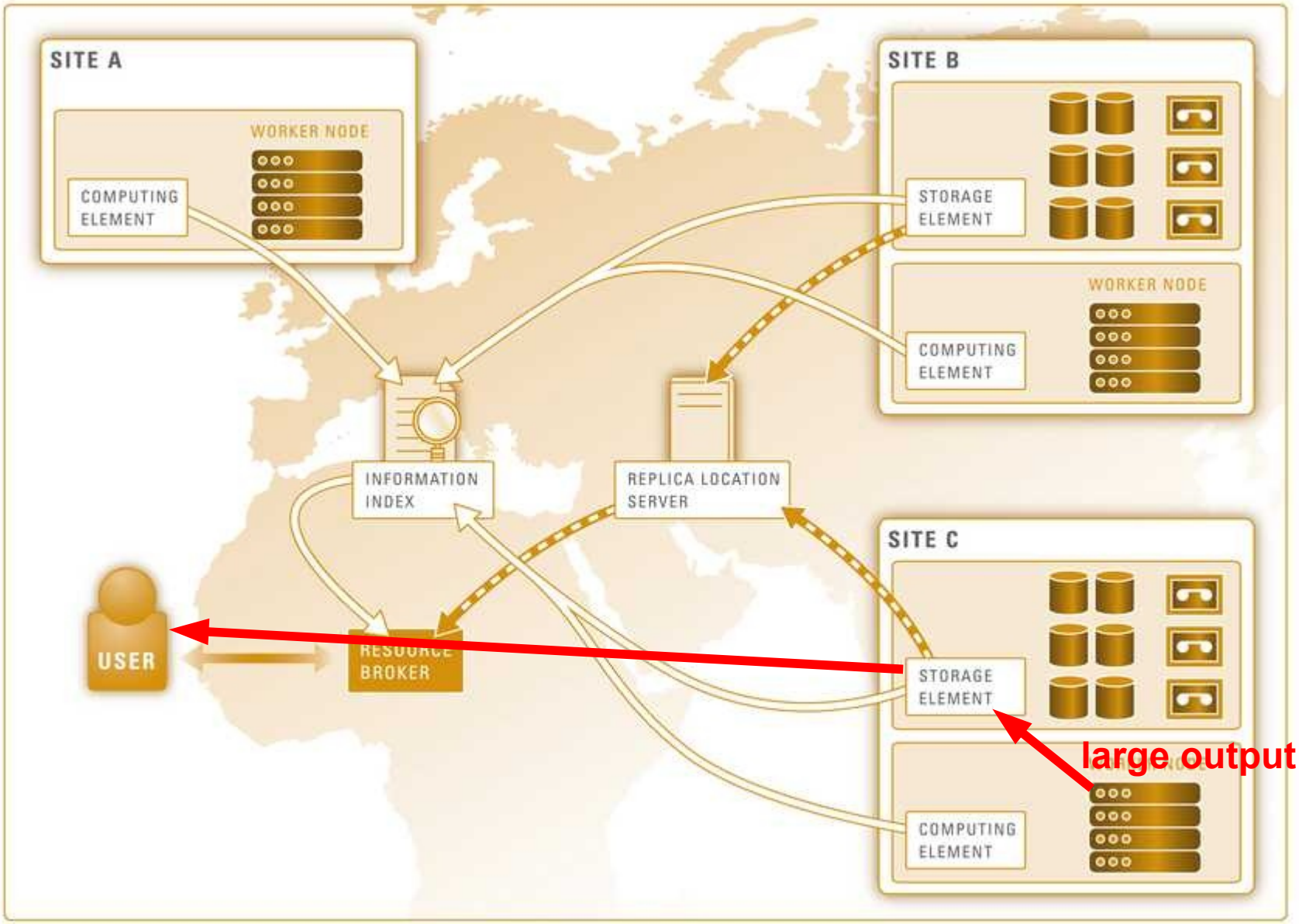
Plot courtesy Martina Hardt / Desginal

Grid Computing



Plot courtesy Martina Hardt / Desginal

Grid Computing



Plot courtesy Martina Hardt / Desginal

EGEE - world's largest Grid infrastructure

- July 2009:
  - ~150 VOs (virtual organisations)
  - ~ 17000 users
  - ~ 290 resource centres in 55 countries
  - ~ 140000 CPUs (cores)
  - > 25 Petabytes of online storage
  - ~330000 jobs per day

- gLite middleware
  - Compute Elemente (CE)
  - Storage Elements (SE)
  - Workload Management Systems (WMS)
  - File Catalogs (LFC)
  - File Transfer Service (FTS)
  - Information System (BDII)
  - ...

http://www.eu-egee.org

http://glite.web.cern.ch

~ 80 people in 12 academic and industrial research centres

EGEE - world's largest Grid infrastructure

- EGEE-III : 2 years
  - EU co-funding ~ 32M€
  - Total budget ~ 47 M€
  - Equipment ~ 50 M€
  - 9132 person months
    - ~ 382 FTE

EGEE activities

Networking

NA1: Project Management
NA2: Dissemination, Communication and Outreach
NA3: User Training and Induction
NA4: User Community Support and Expansion
NA5: Policy and International Cooperation

Services

SA1: Grid Operations, Support and Management
SA2: Networking Support
SA3: Integration, Testing and Certification

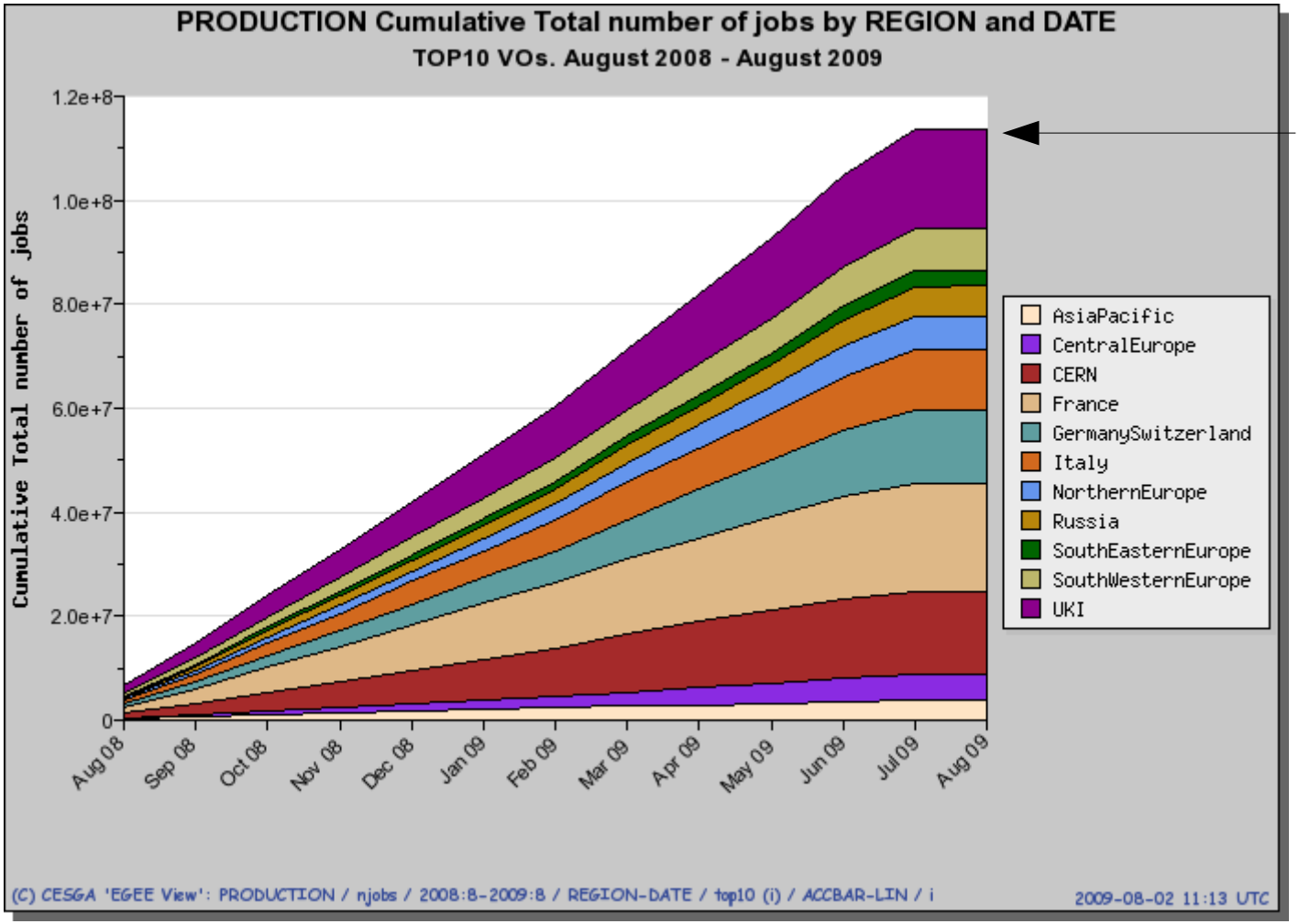
Joint Research

JRA1: Middleware Re-engineering

EGEE - world's largest Grid infrastructure

EGEE - world's largest Grid infrastructure

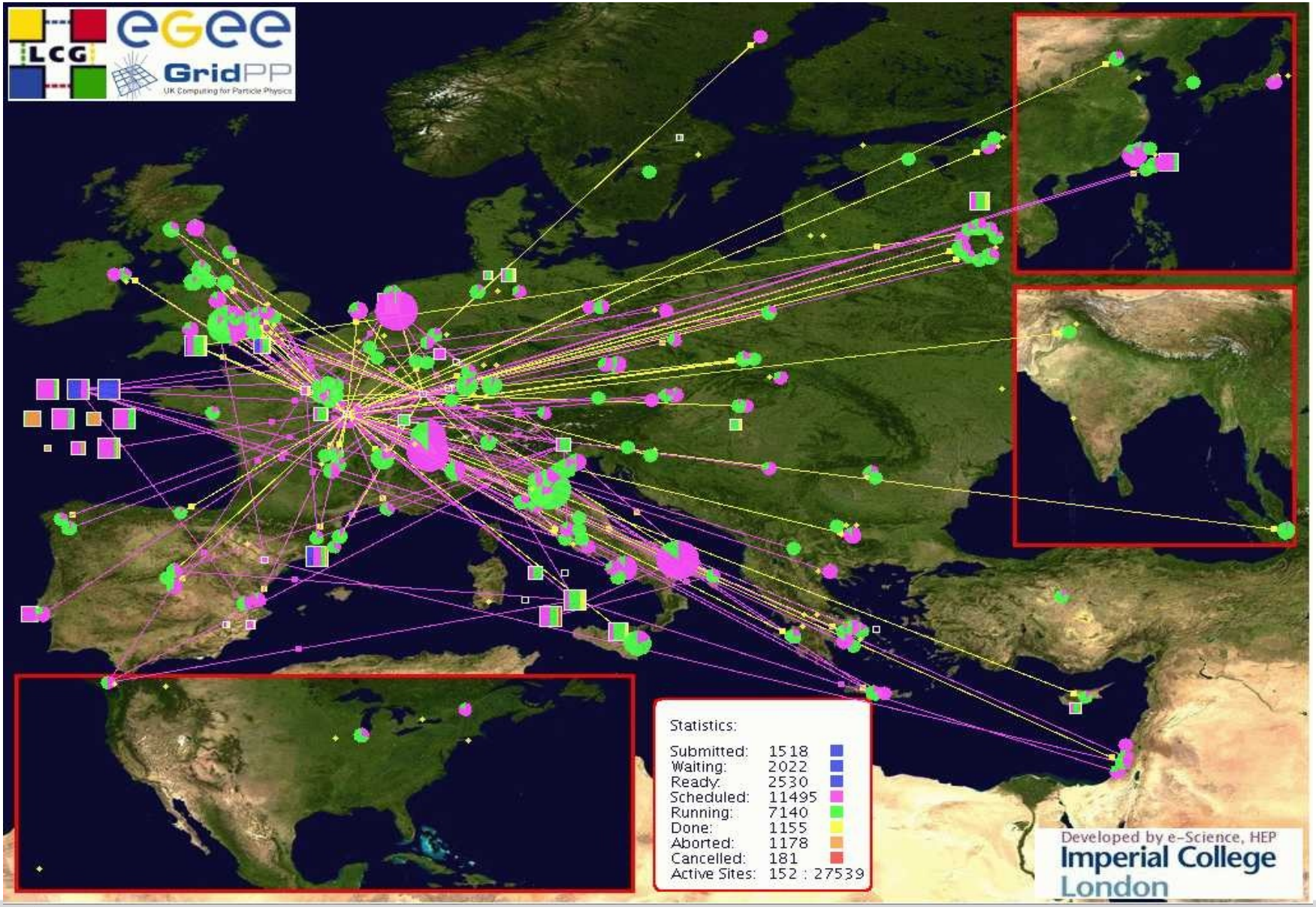
no. of jobs in EGEE last 12 months



110,000,000

EGEE - world's largest Grid infrastructure

http://gridportal.hep.ph.ic.ac.uk/rtm/

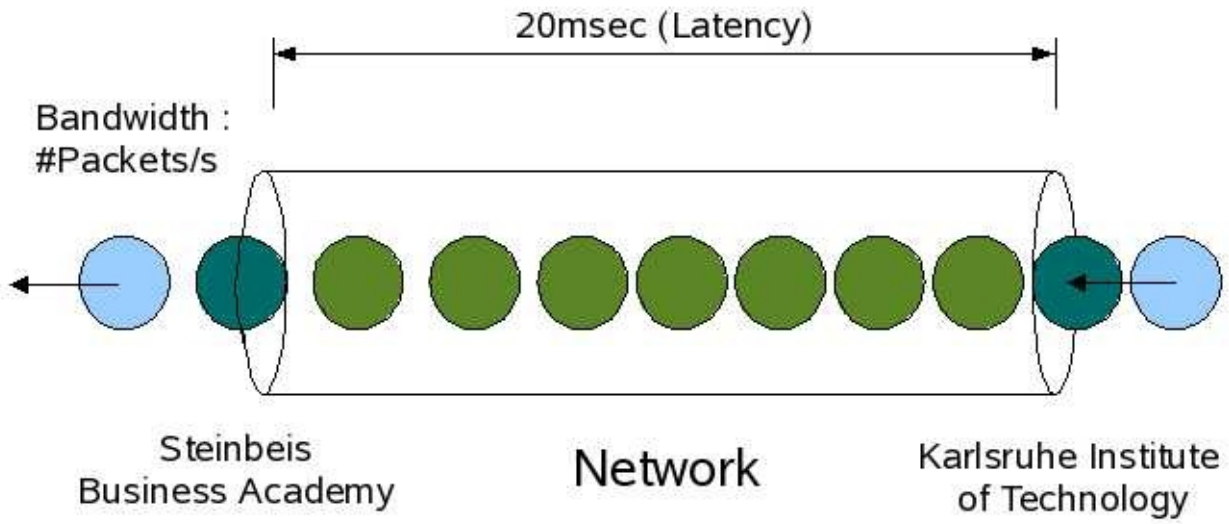


Bandwidth vs. latency

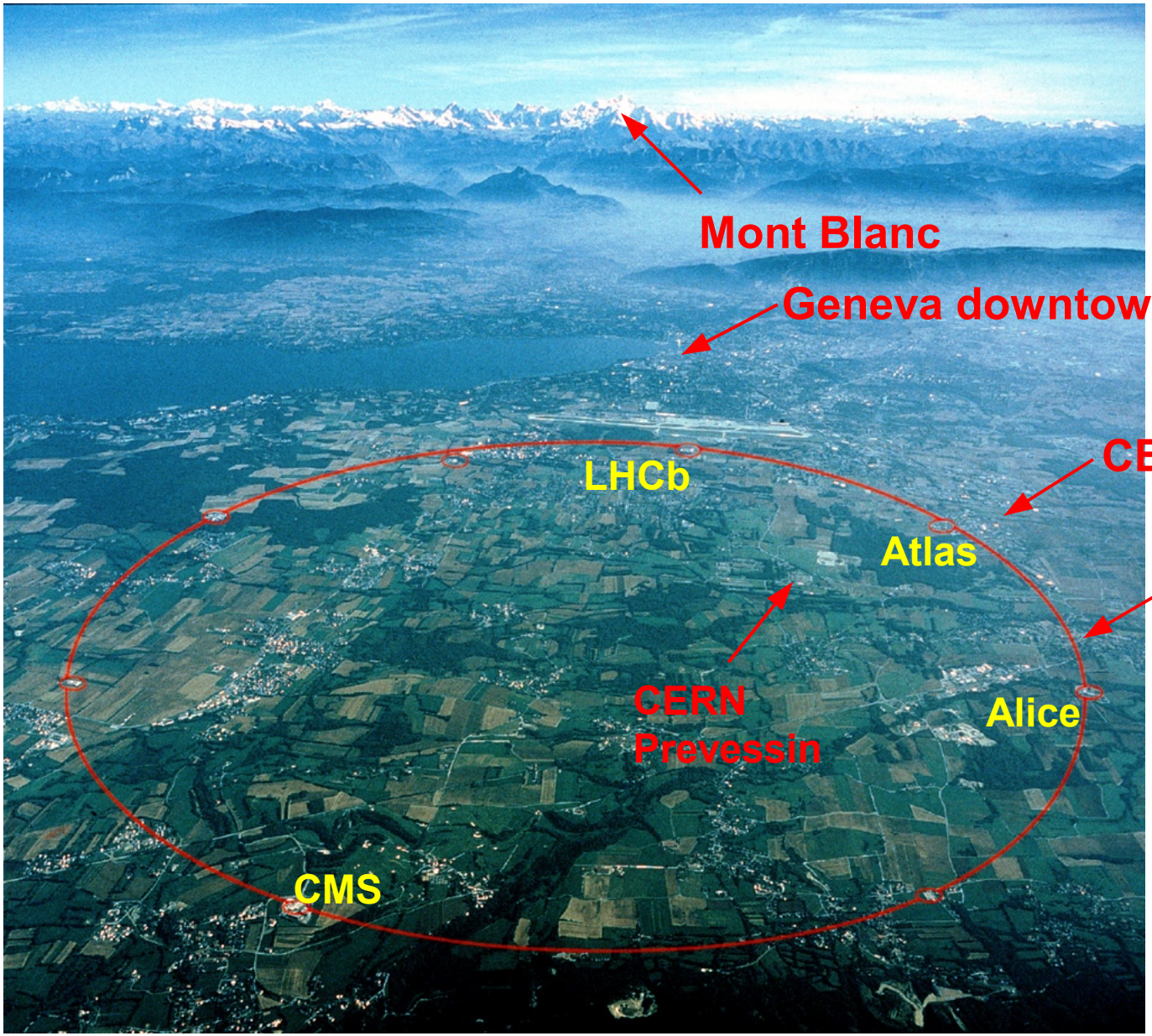
- Network bandwidth scales almost indefinite (matter of money)
- Latency does not scale!

→ Not all computational problems are suited for the Grid (parallel computing: e.g. weather simulation / forecast)

→ Grids ideally suited for embarrassingly parallel workloads, small and large data volumes. (e.g. MC simulation and data analysis for (astro)particle physics.)

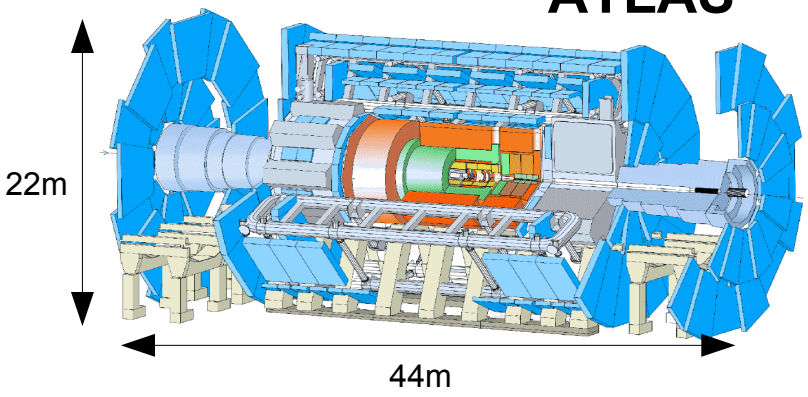


The Large Hadron Collider (LHC)

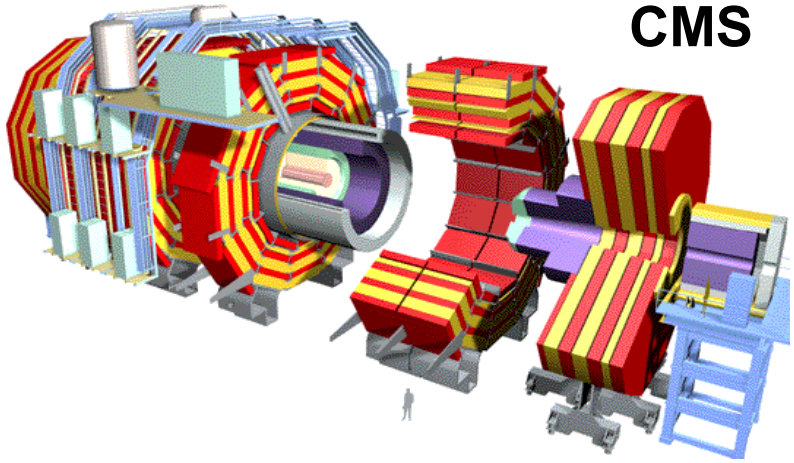


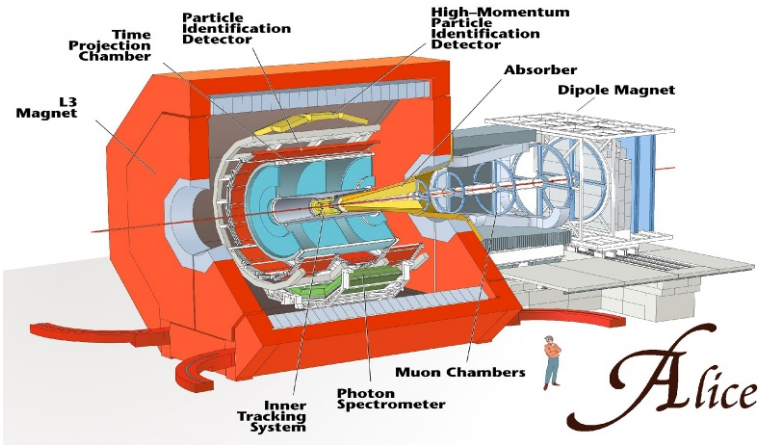
The Large Hadron Collider (LHC)

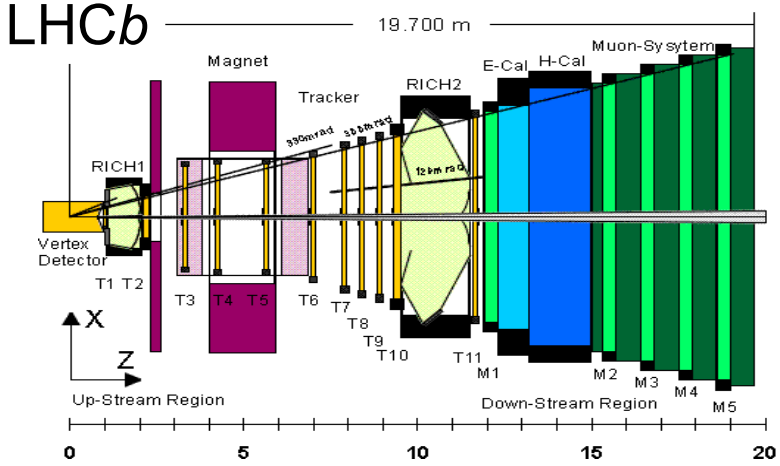
ATLAS



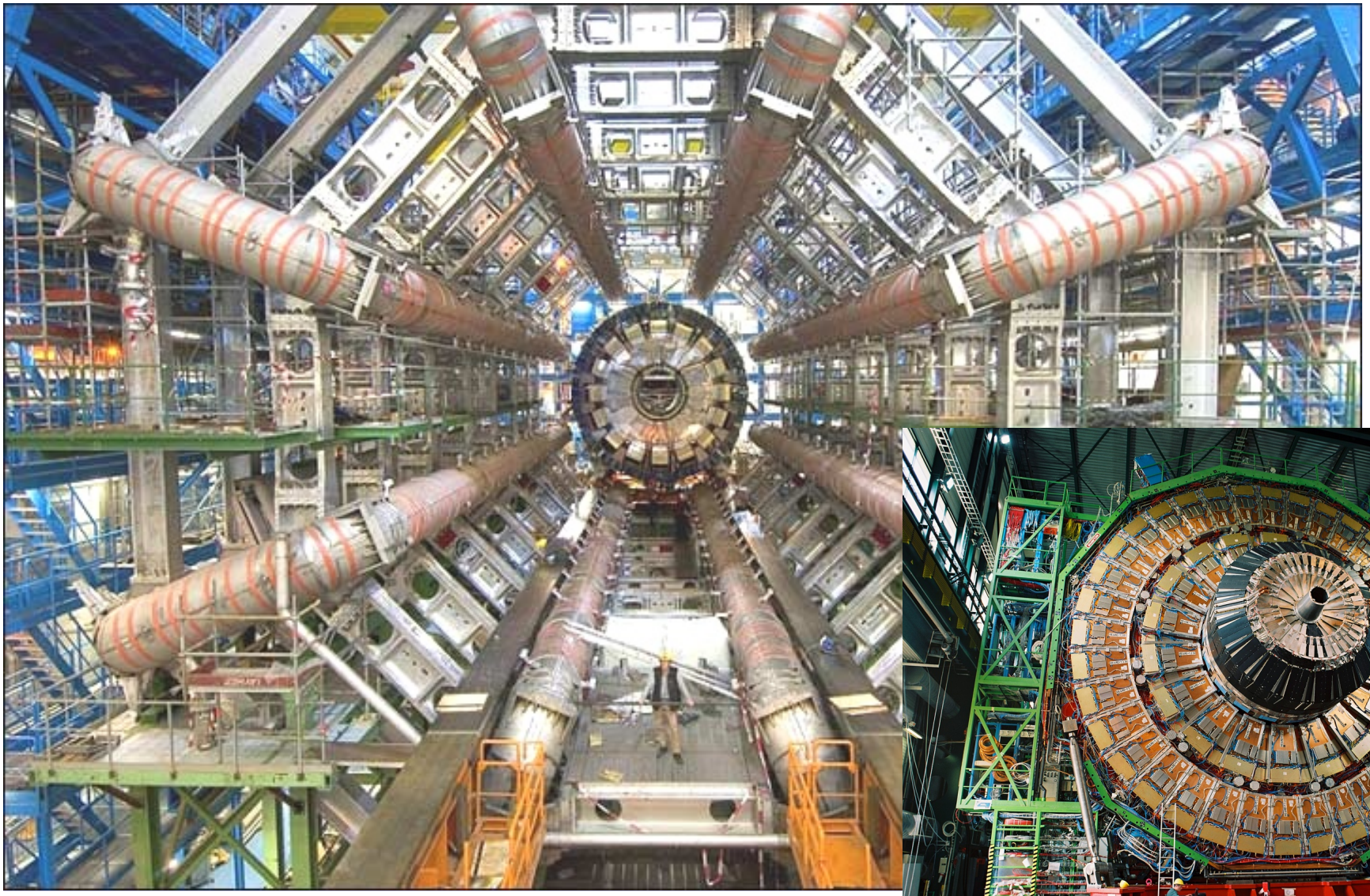
CMS



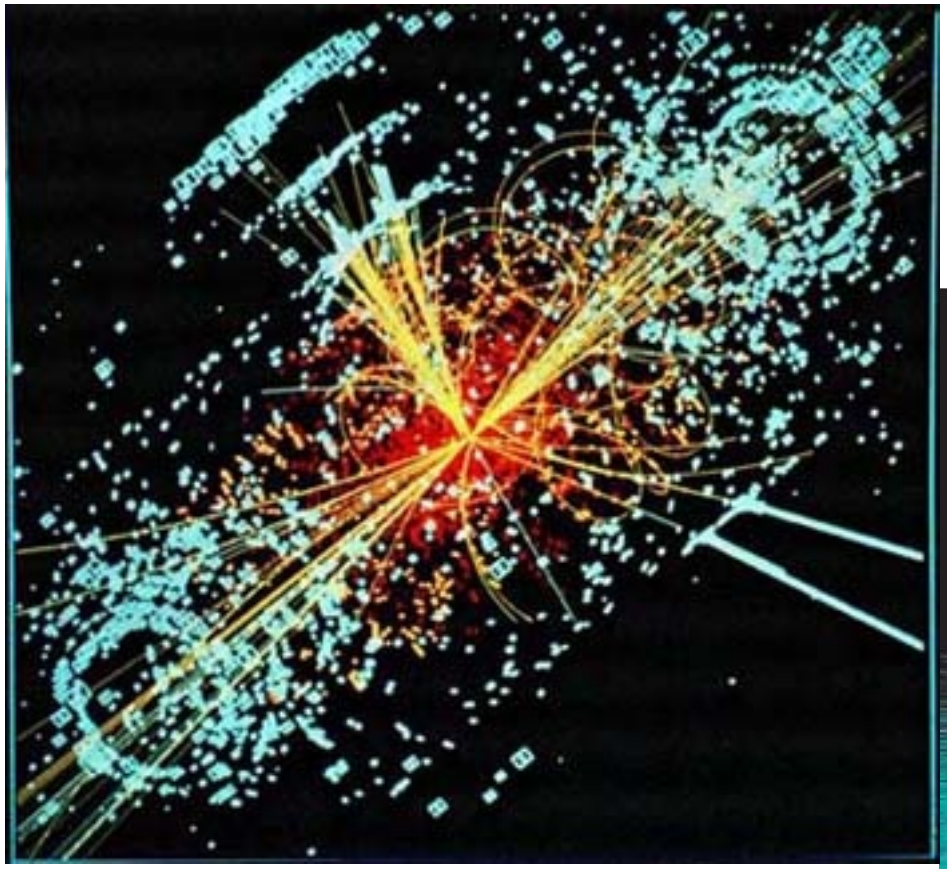




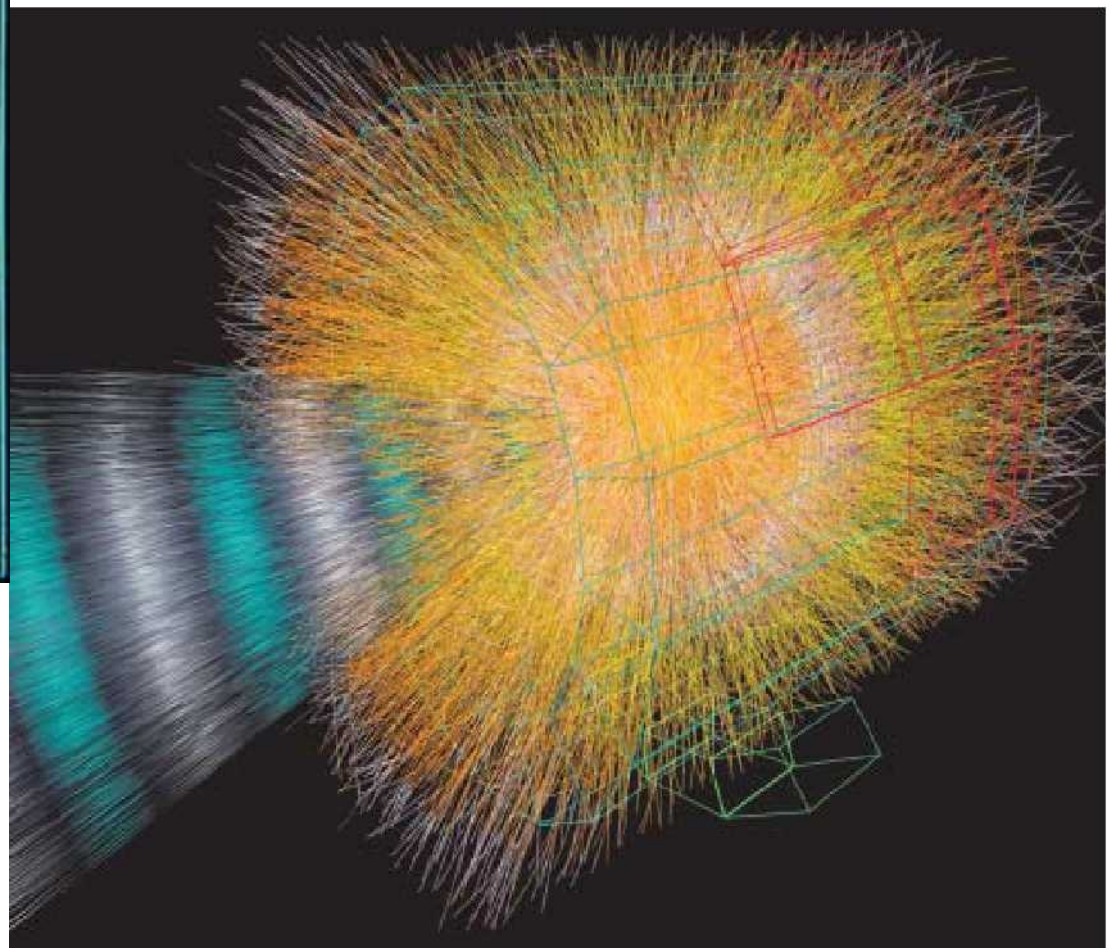
The Large Hadron Collider (LHC)



The Large Hadron Collider (LHC)

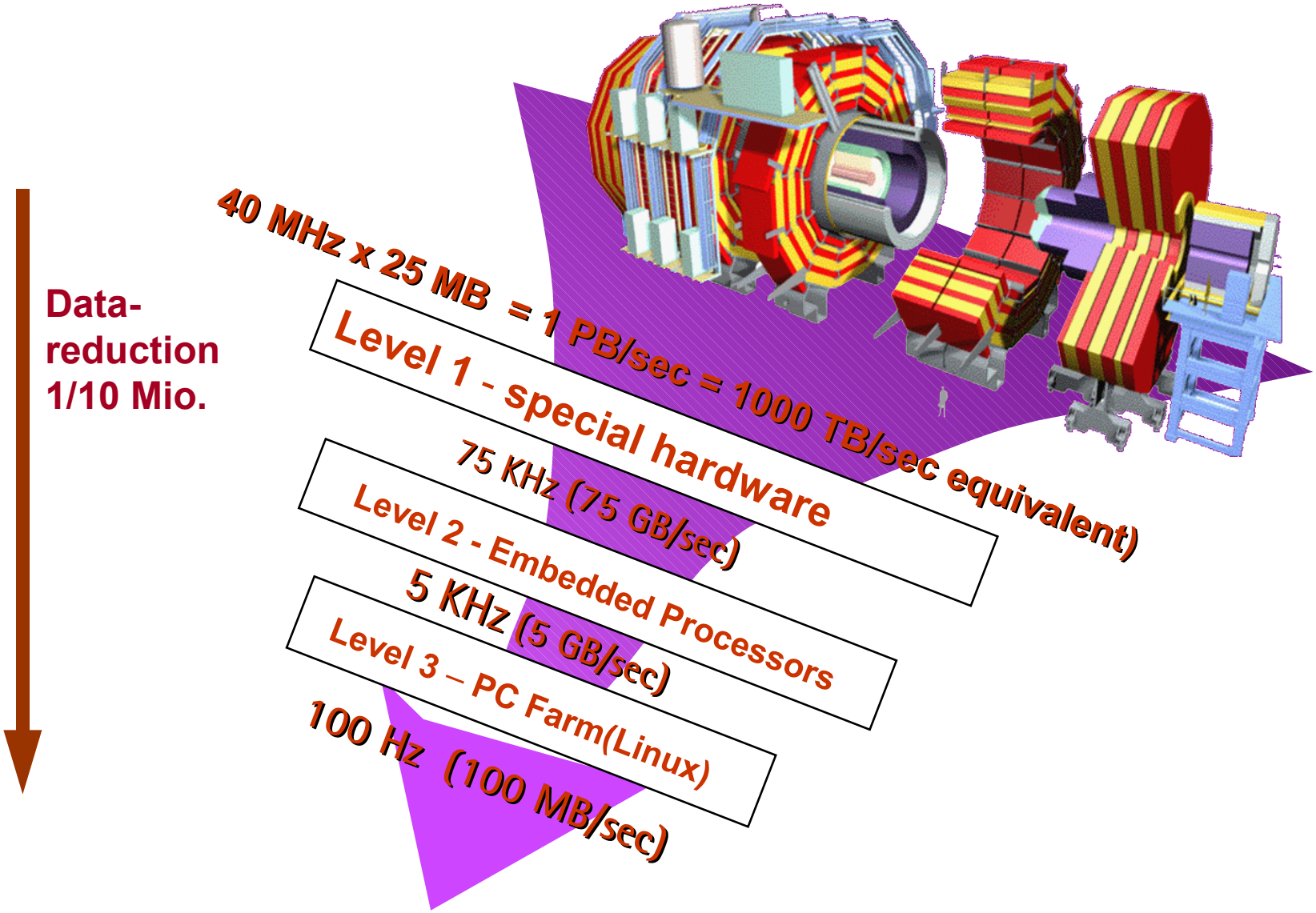


Tracks from a Higgs decay in the
CMS tracker (76 mio. readout channels)



Alice:
40 MHz collisions of lead nuclei

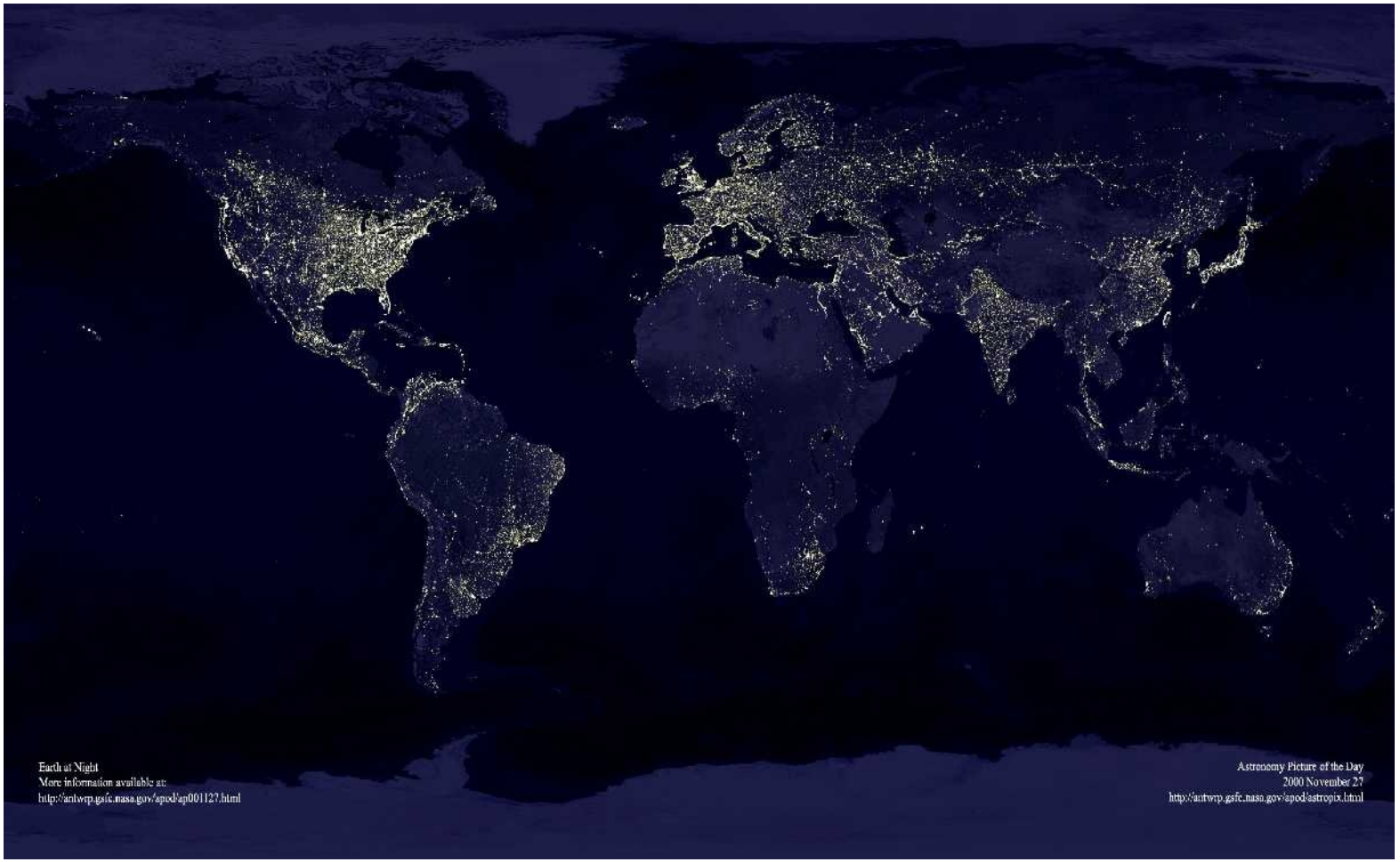
The Worldwide LHC Computing Grid (WLCG)



~ 1-2 PB per year per experiment (+ MC data)

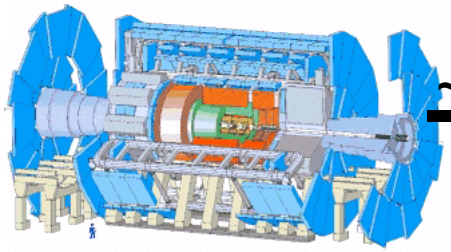
The Worldwide LHC Computing Grid (WLCG)

10000 physicists worldwide ...



.... want to analyse the LHC data.

The Worldwide LHC Computing Grid (WLCG)



~PB/s

Online System

100 - 400 MB/s per Exp.

Offline farm
CERN computer centre

Tier-1

Tier-1

GridKa

• • •

reconstructed data
MC

DESY

Tier-2

Tier-2

• • •

Institute

Institu

Institu

Institute

Tier-3

10000 physicists worldwide

Tier-1 (11 sites)

- Manage permanent storage (RAW, simulated, processed)
- Capacity for re-processing and bulk analysis

Tier-2 (~120 sites)

- Monte Carlo event simulation
- User analysis

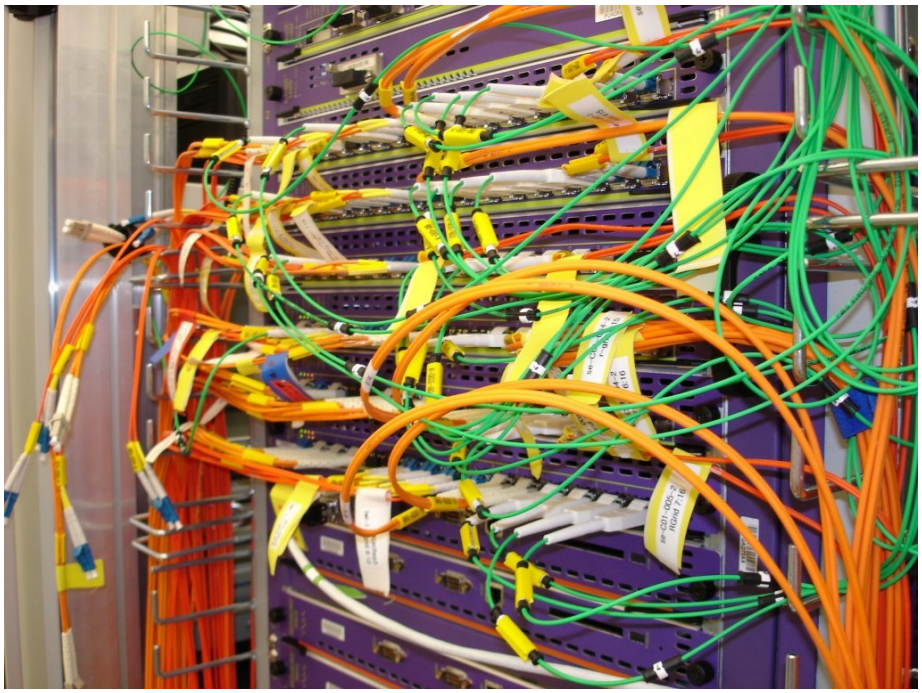
GridKa - the German WLCG Tier-1

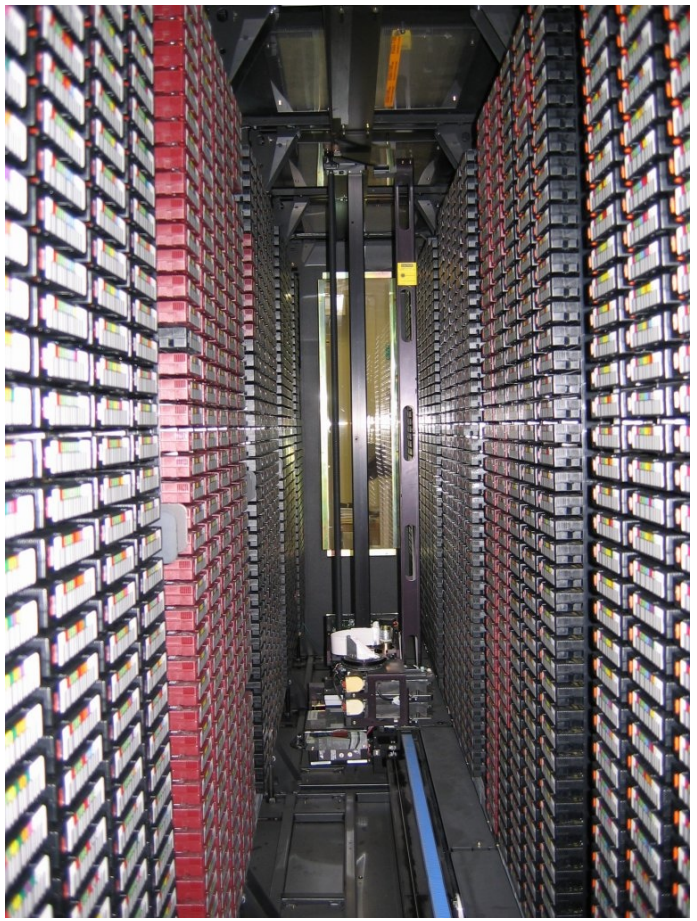
- Started 2002, requested by the German High Energy Physics community
- Supports many international experiments
  - LHC: all 4 experiments, Alice, Atlas, CMS, LHCb
  - non-LHC HEP: Babar, (Super)Belle, CDF, Compass, D0
  - Astroparticle, other non-HEP: Auger, Magic, Medigrid, ...
- Provides Resources and services for EGEE and D-Grid

- 7820 CPU cores
- ~ 30000 jobs per day

GridKa - the German WLCG Tier-1

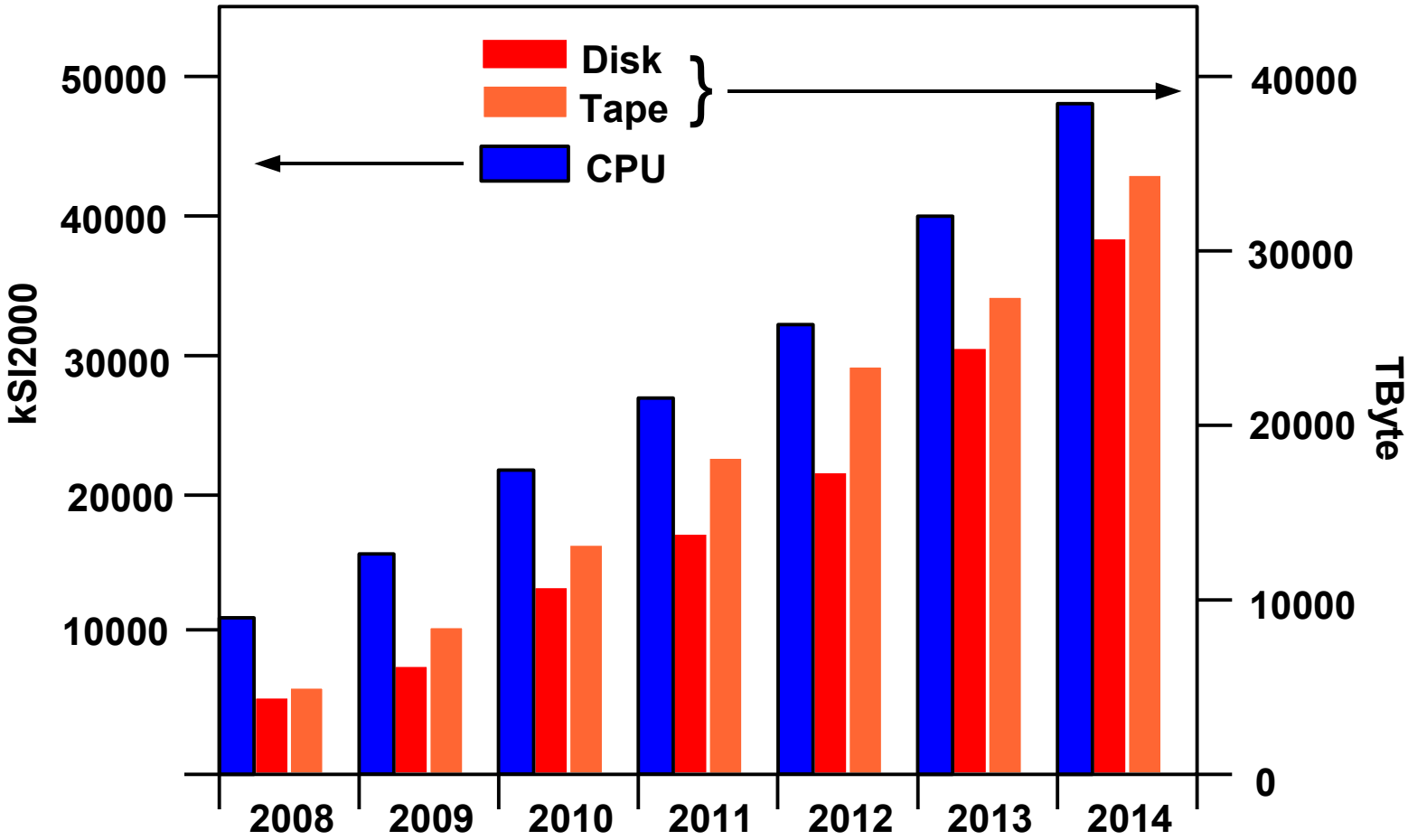
- 6300 TB disk
- 8500 TB tape
- > 50 Gbit/s WAN



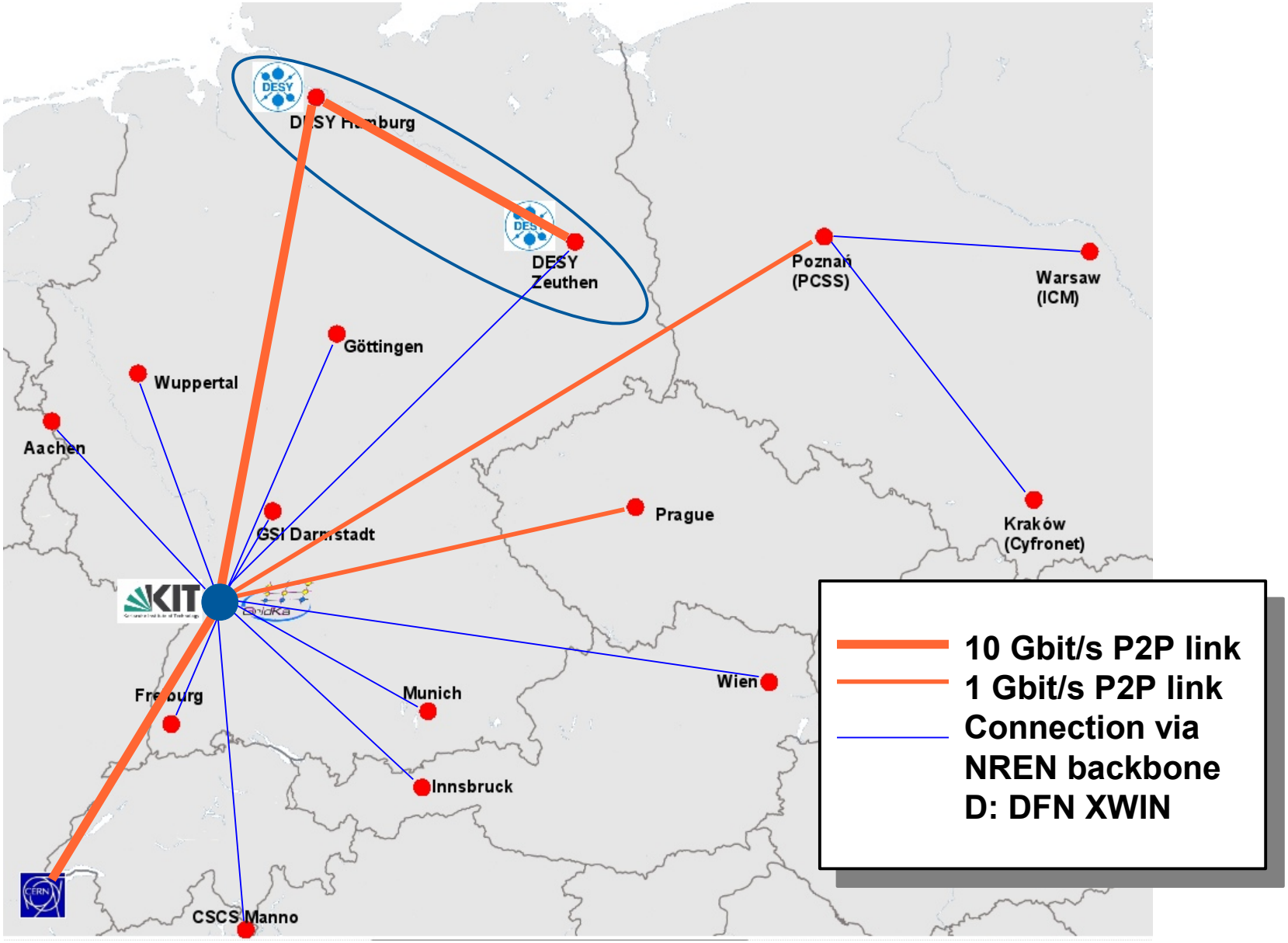


GridKa - the German WLCG Tier-1

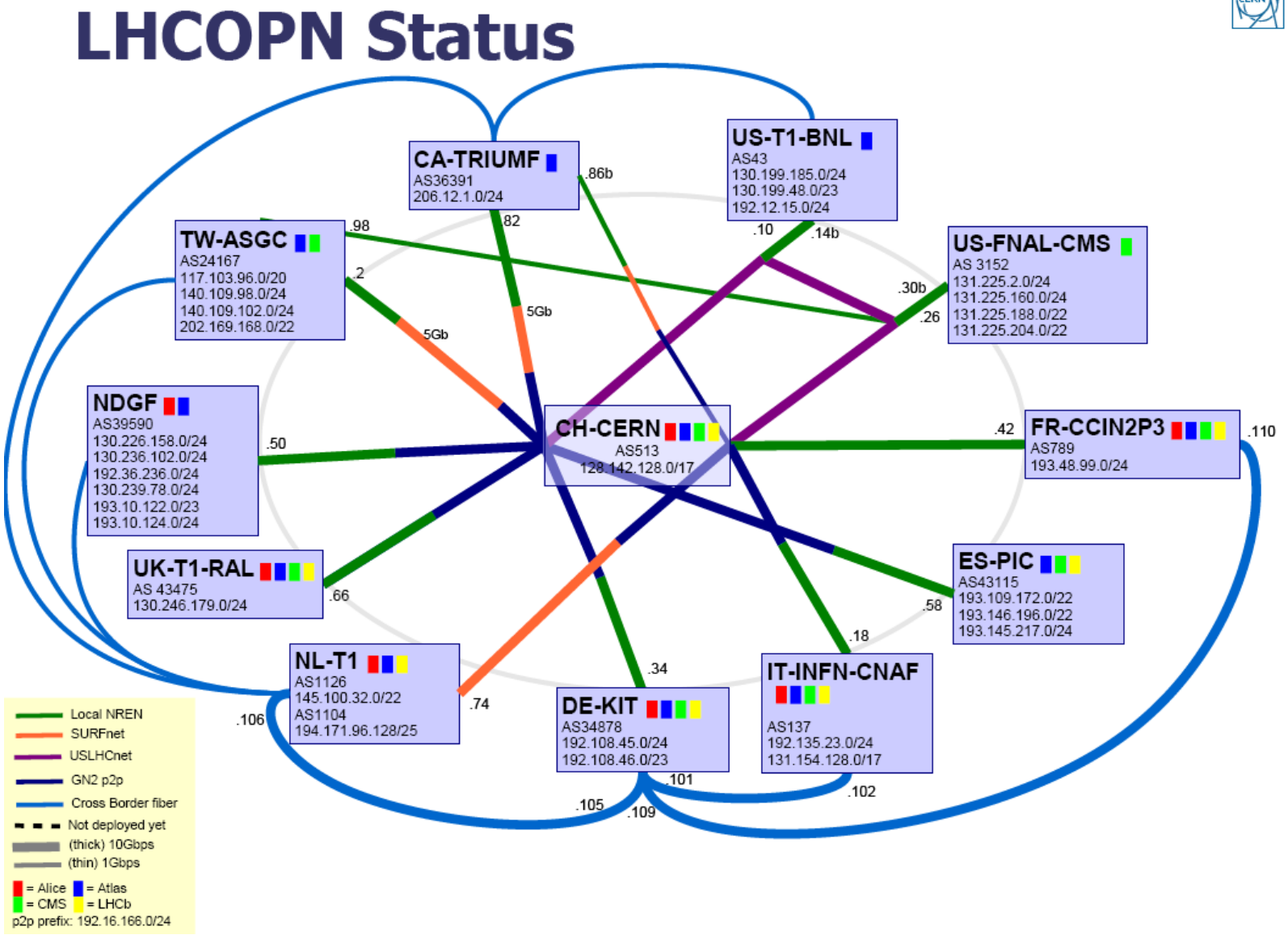
- Storage and computing resources



GridKa and associated Tier-2 sites

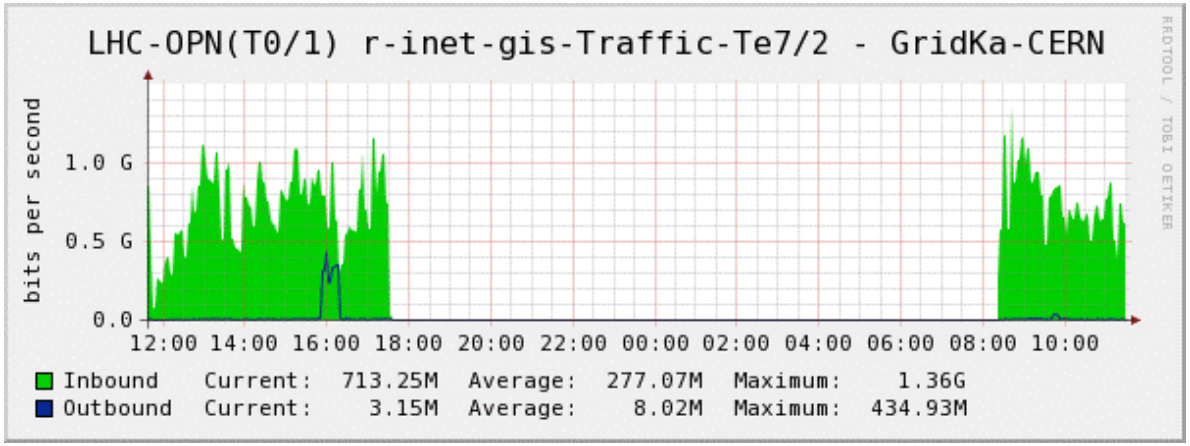


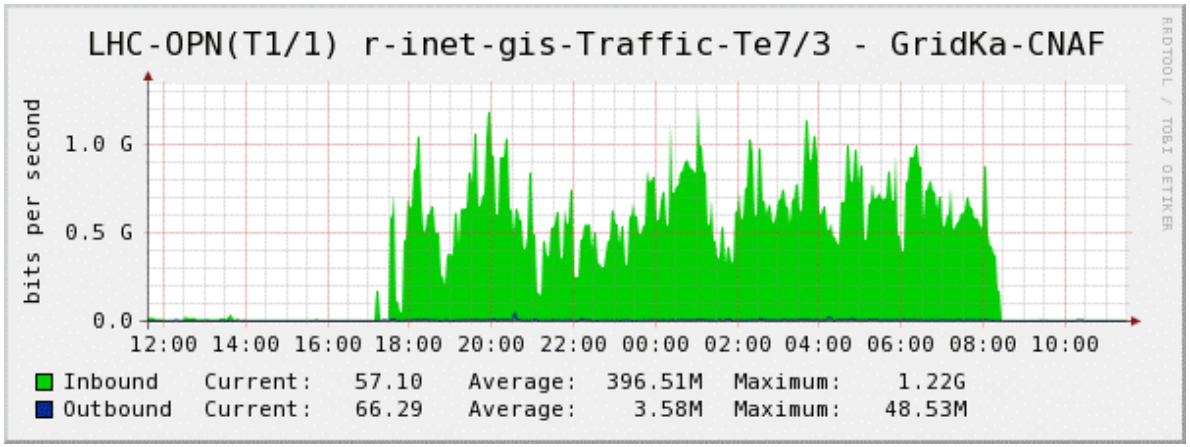
GridKa WAN connections



GridKa WAN connections

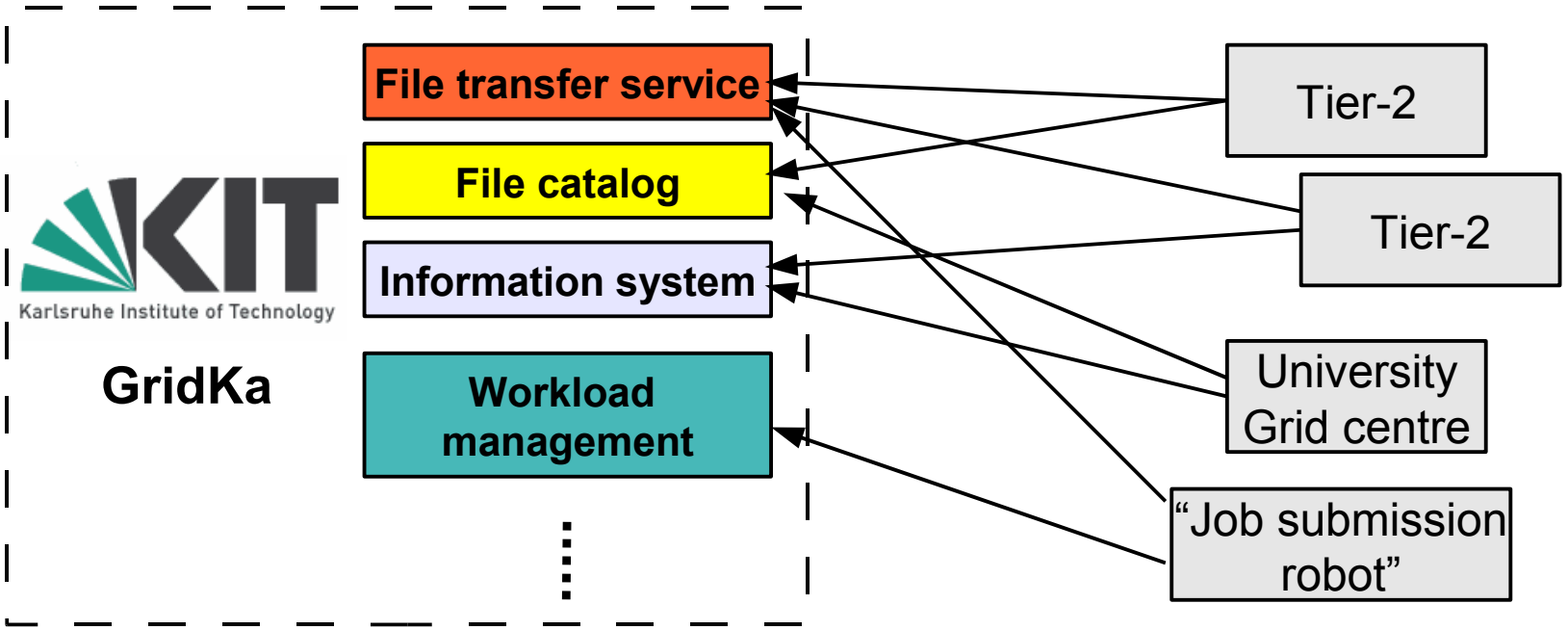
Network failure of the LHCOPN link between CERN and FZK on April 26th / 27th 2007





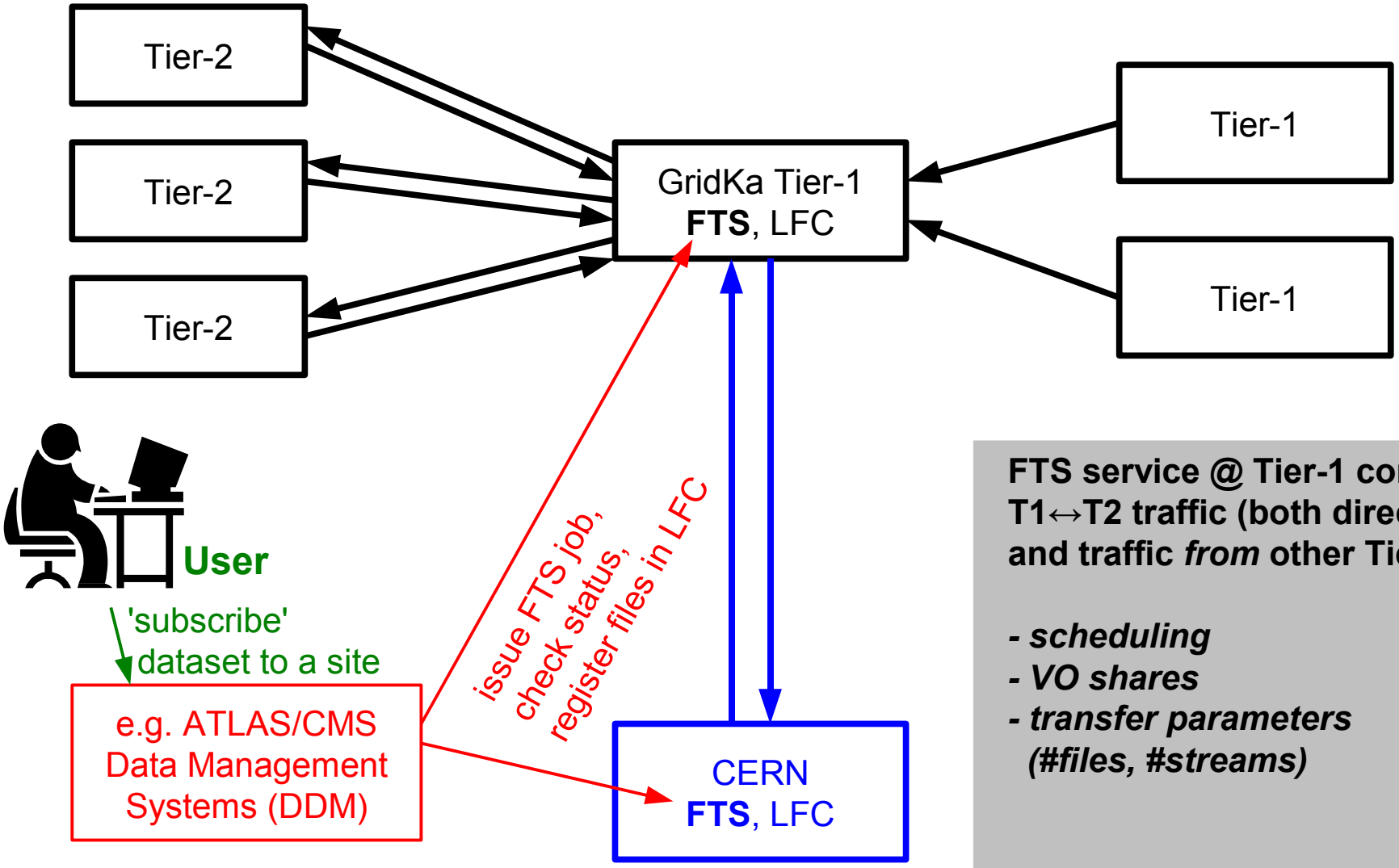
automatic routing of T0-T1 traffic over the backup link via CNAF

GridKa services for WLCG and EGEE (regional core services)



- Used by Tier-2, university Grid centres and experiment-specific high level services (job submission robots, data management systems)
- Highest reliability necessary

Data management in WLCG



FTS service @ Tier-1 controls T1↔T2 traffic (both directions) and traffic *from* other Tier-1s.

- ***scheduling***
- ***VO shares***
- ***transfer parameters (#files, #streams)***

OVERVIEW

Activity Period

Activity in Last Hour
Activity in Last 4 Hours
Activity in Last 24 Hours
Activity in Last 7 Days
Activity in Last 30 Days
Activity in ...

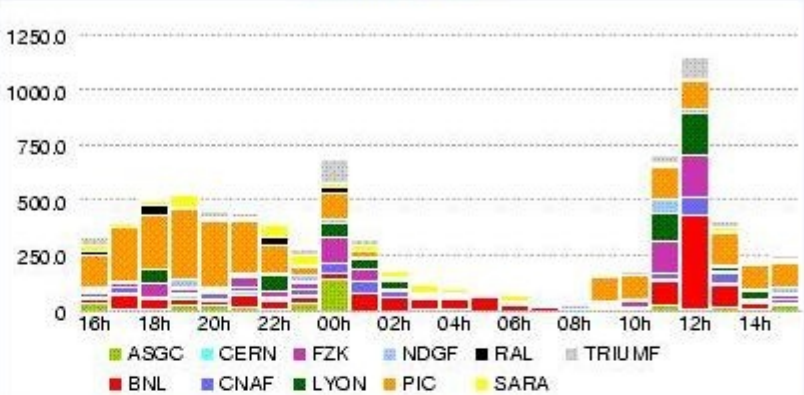
Selected Activities

Production
T0 Export
Functional Test
User Subscriptions
Staging
Data Consolidation

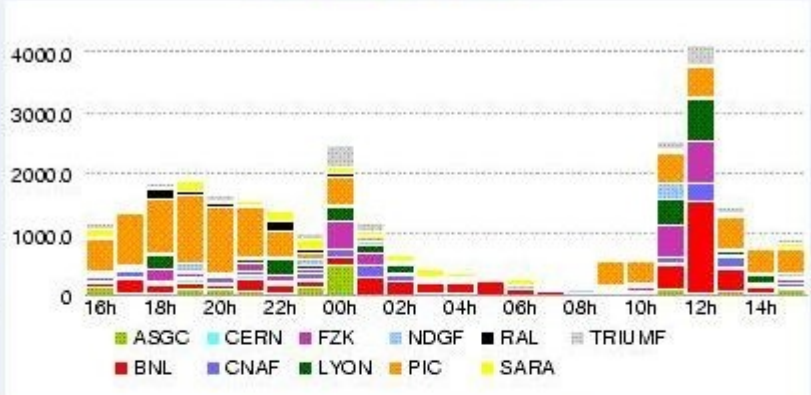
Selected Cloud

ASGC Cloud
BNL Cloud
CERN Cloud
CNAF Cloud
FZK Cloud
LYON Cloud
NDGF Cloud
PIC Cloud
RAL Cloud
SARA Cloud
TRIUMF Cloud

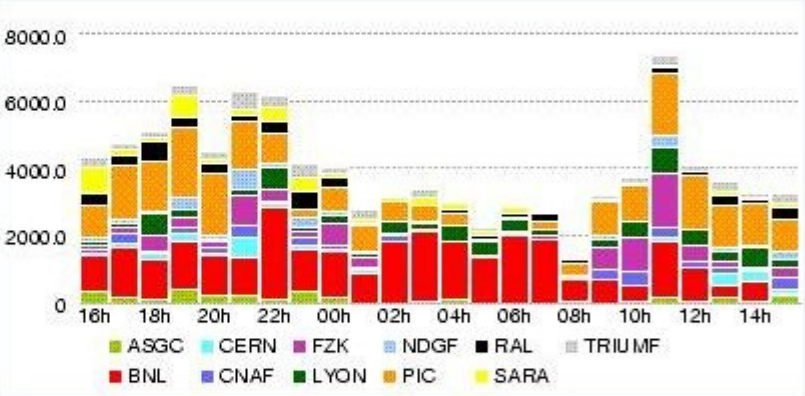
Throughput (MB/s)



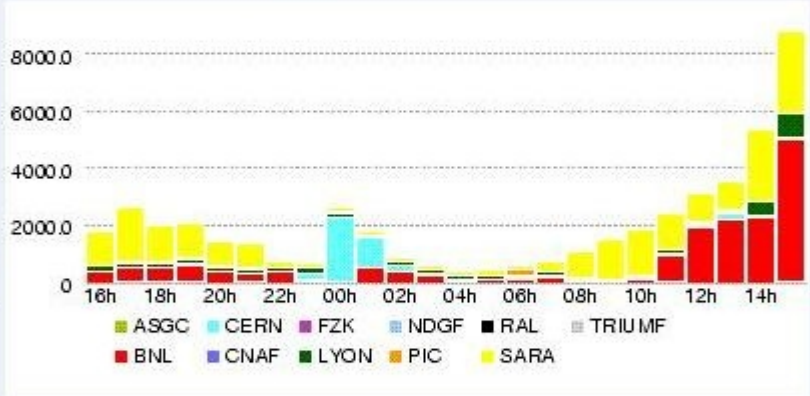
Data Transferred (GBytes)



Completed File Transfers



Total Number Transfer Errors

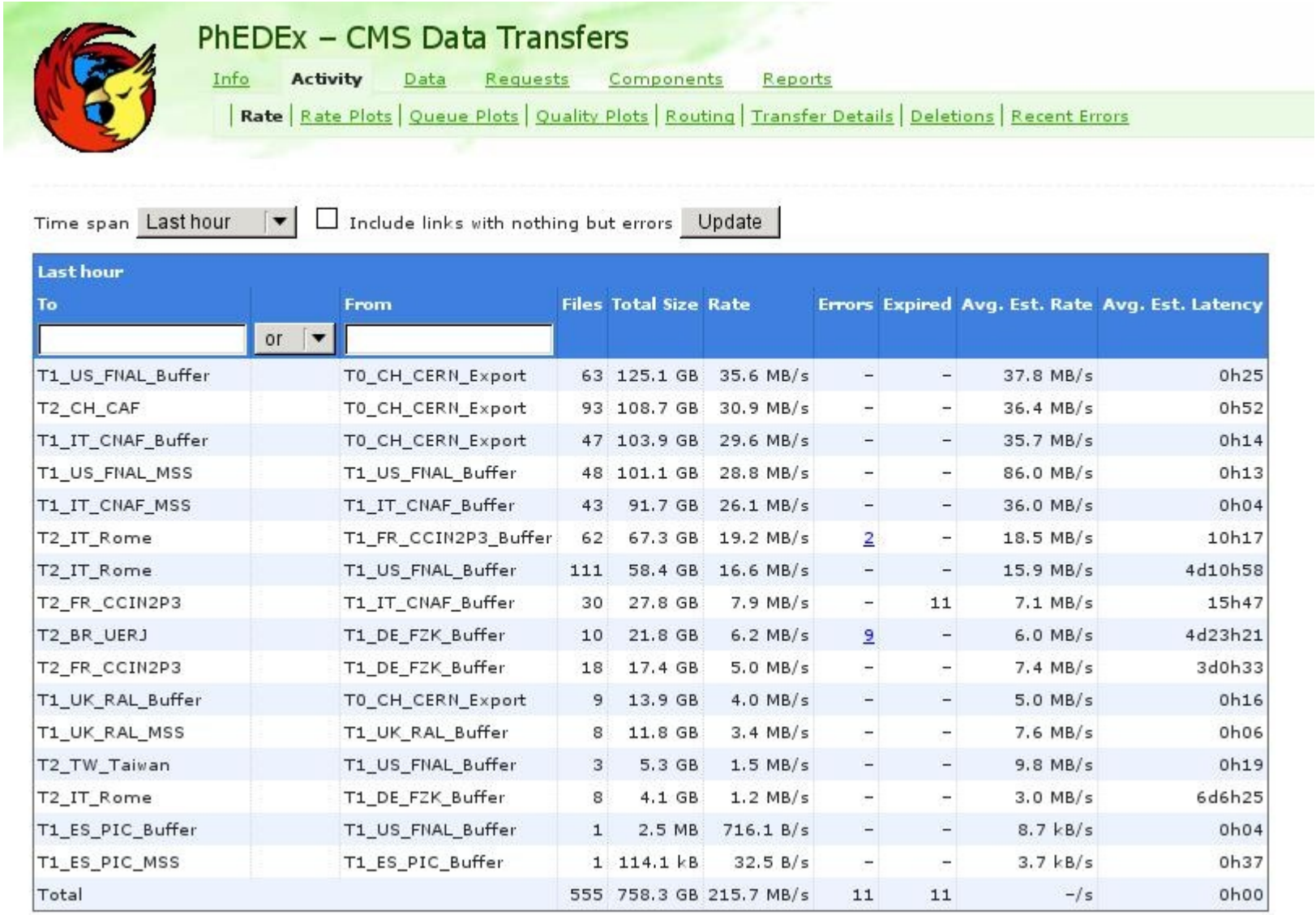


Activity Summary ('2009-07-27 16:00' to '2009-07-28 16:00')

Click on the cloud name to view list of sites

	Transfers			Registrations		Errors			Services
Cloud	Efficiency	Throughput	Successes	Datasets	Files	Transfer	Registration	Services	Grid
ASGC	99%	14 MB/s	2627	1402	2618	19	0	0	
BNL	63%	56 MB/s	30439	3172	30683	18246	0	0	
CERN	35%	5 MB/s	2514	715	2514	4614	0	0	
CNAF	98%	20 MB/s	3804	1478	3790	74	0	0	
FZK	98%	33 MB/s	8513	2469	8507	167	0	0	
LYON	64%	31 MB/s	7455	2537	7449	4136	0	0	
NDGF	99%	15 MB/s	3012	1349	3016	43	0	0	
PIC	97%	112 MB/s	23564	1538	23554	658	0	0	

CMS data management



PhEDEx – CMS Data Transfers

Info | **Activity** | Data | Requests | Components | Reports

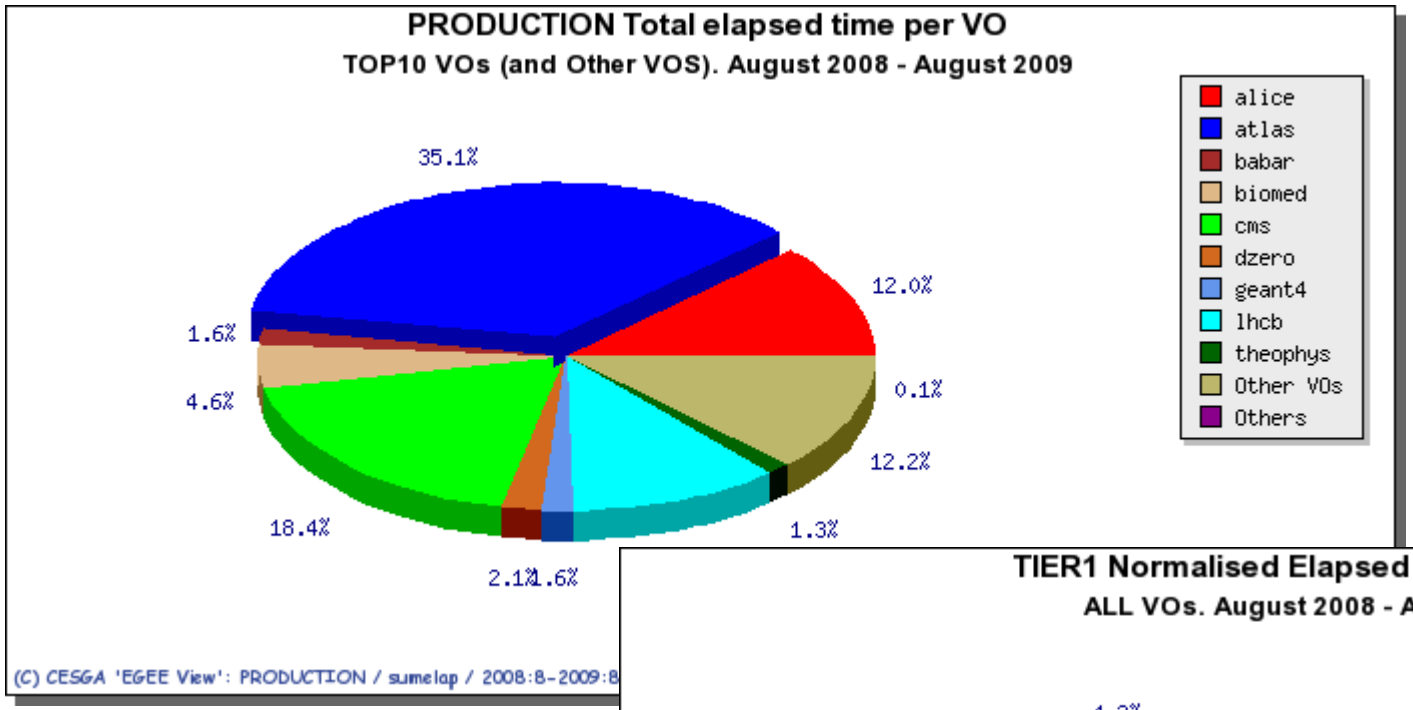
| **Rate** | Rate Plots | Queue Plots | Quality Plots | Routing | Transfer Details | Deletions | Recent Errors

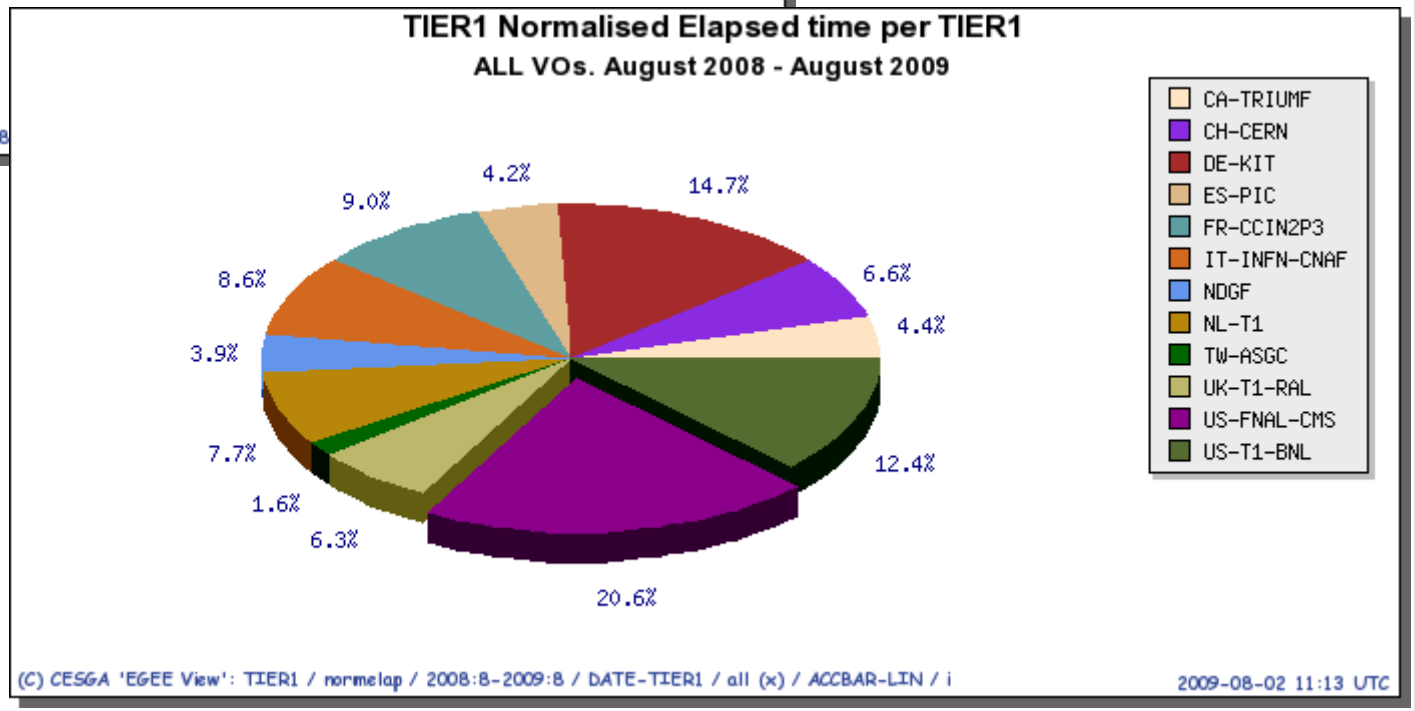
Time span: Last hour ☐ Include links with nothing but errors Update

Last hour

To		From	Files	Total Size	Rate	Errors	Expired	Avg. Est. Rate	Avg. Est. Latency
	or								
T1_US_FNAL_Buffer		T0_CH_CERN_Export	63	125.1 GB	35.6 MB/s	-	-	37.8 MB/s	0h25
T2_CH_CAF		T0_CH_CERN_Export	93	108.7 GB	30.9 MB/s	-	-	36.4 MB/s	0h52
T1_IT_CNAF_Buffer		T0_CH_CERN_Export	47	103.9 GB	29.6 MB/s	-	-	35.7 MB/s	0h14
T1_US_FNAL_MSS		T1_US_FNAL_Buffer	48	101.1 GB	28.8 MB/s	-	-	86.0 MB/s	0h13
T1_IT_CNAF_MSS		T1_IT_CNAF_Buffer	43	91.7 GB	26.1 MB/s	-	-	36.0 MB/s	0h04
T2_IT_Rome		T1_FR_CCIN2P3_Buffer	62	67.3 GB	19.2 MB/s	2	-	18.5 MB/s	10h17
T2_IT_Rome		T1_US_FNAL_Buffer	111	58.4 GB	16.6 MB/s	-	-	15.9 MB/s	4d10h58
T2_FR_CCIN2P3		T1_IT_CNAF_Buffer	30	27.8 GB	7.9 MB/s	-	11	7.1 MB/s	15h47
T2_BR_UERJ		T1_DE_FZK_Buffer	10	21.8 GB	6.2 MB/s	9	-	6.0 MB/s	4d23h21
T2_FR_CCIN2P3		T1_DE_FZK_Buffer	18	17.4 GB	5.0 MB/s	-	-	7.4 MB/s	3d0h33
T1_UK_RAL_Buffer		T0_CH_CERN_Export	9	13.9 GB	4.0 MB/s	-	-	5.0 MB/s	0h16
T1_UK_RAL_MSS		T1_UK_RAL_Buffer	8	11.8 GB	3.4 MB/s	-	-	7.6 MB/s	0h06
T2_TW_Taiwan		T1_US_FNAL_Buffer	3	5.3 GB	1.5 MB/s	-	-	9.8 MB/s	0h19
T2_IT_Rome		T1_DE_FZK_Buffer	8	4.1 GB	1.2 MB/s	-	-	3.0 MB/s	6d6h25
T1_ES_PIC_Buffer		T1_US_FNAL_Buffer	1	2.5 MB	716.1 B/s	-	-	8.7 kB/s	0h04
T1_ES_PIC_MSS		T1_ES_PIC_Buffer	1	114.1 kB	32.5 B/s	-	-	3.7 kB/s	0h37
Total			555	758.3 GB	215.7 MB/s	11	11	-/s	0h00

EGEE jobs last year





A Grid job example (1)

```
heiss@gridka25:~> voms-proxy-init  --voms dteam
Enter GRID pass phrase:
Your identity: /O=GermanGrid/OU=FZK/CN=Andreas Heiss
Creating temporary proxy ..................................... Done
Contacting  voms.cern.ch:15004 [/DC=ch/DC=cern/OU=computers/CN=voms.cern.ch]
"dteam" Done
Creating proxy ................................................ Done
Your proxy is valid until Fri Jul 24 02:11:55 2009
```

```
heiss@gridka25:~> cat job.jdl
Type = "Job";
JobType = "Normal";
Executable = "/bin/hostname";
StdOutput = "hello.out";
StdError = "hello.err";
OutputSandbox = {"hello.err","hello.out"};
RetryCount = 2;
VirtualOrganisation = "dteam";
```

A Grid job example (2)

```
heiss@gridka25:~> glite-wms-job-list-match -a job.jdl

Connecting to the service https://wms-3-fzk.gridka.de:7443/glite_wms_wmproxy_server

================================================================================

                     COMPUTING ELEMENT IDs LIST
 The following CE(s) matching your job requirements have been found:

        *CEId*
 - agh2.atlas.unimelb.edu.au:2119/jobmanager-lcgpbs-dteam
 - alice003.nipne.ro:2119/jobmanager-lcgpbs-dteam
 - alice19.spbu.ru:2119/jobmanager-lcgpbs-dteam
 - atlasce.phys.sinica.edu.tw:2119/jobmanager-lcgcondor-dteam
 - atlasce01.na.infn.it:2119/jobmanager-lcgpbs-cert
 - axon-g01.ieeta.pt:2119/jobmanager-lcgpbs-dteam
 - bigmac-lcg-ce2.physics.utoronto.ca:2119/jobmanager-pbs-dteam
 - bugaboo-hep.westgrid.ca:2119/jobmanager-lcgpbs-dteam
 - ce-01.grid.sissa.it:2119/jobmanager-lcgpbs-cert
 - ce-01.roma3.infn.it:2119/jobmanager-lcgpbs-cert
 - ce-1-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
 - ce-2-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
 - ce-3-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
 - ce-4-fzk.gridka.de:2119/jobmanager-pbspro-gLite3
 - ce-alice.sdfarm.kr:2119/jobmanager-lcgpbs-dteam
 - ce-cyb.ca.infn.it:2119/jobmanager-lcglsf-poncert
 - ......
```

A Grid job example (3)

```
heiss@gridka25:~> glite-wms-job-submit -a job.jdl

Connecting to the service https://wms-3-fzk.gridka.de:7443/glite_wms_wmproxy_server


====================== glite-wms-job-submit Success ======================

The job has been successfully submitted to the WMProxy
Your job identifier is:

https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw

==========================================================================
```

A Grid job example (4)

```
heiss@gridka25:~> glite-wms-job-status https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyI..

**************************************************************
BOOKKEEPING INFORMATION:

Status info for the Job : https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw
Current Status:     Scheduled
Status Reason:      Job successfully submitted to Globus
Destination:        grid-ce01.esrf.eu:2119/jobmanager-pbs-short
Submitted:          Thu Jul 23 14:50:49 2009 CEST
**************************************************************


heiss@gridka25:~> glite-wms-job-status https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uy..
**************************************************************
BOOKKEEPING INFORMATION:

Status info for the Job : https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw
Current Status:     Done (Success)
Logged Reason(s):
    -
    - Job terminated successfully
Exit code:          0
Status Reason:      Job terminated successfully
Destination:        grid-ce01.esrf.eu:2119/jobmanager-pbs-short
Submitted:          Thu Jul 23 14:50:49 2009 CEST
**************************************************************
```

European synchrotron rad. facility, Grenoble

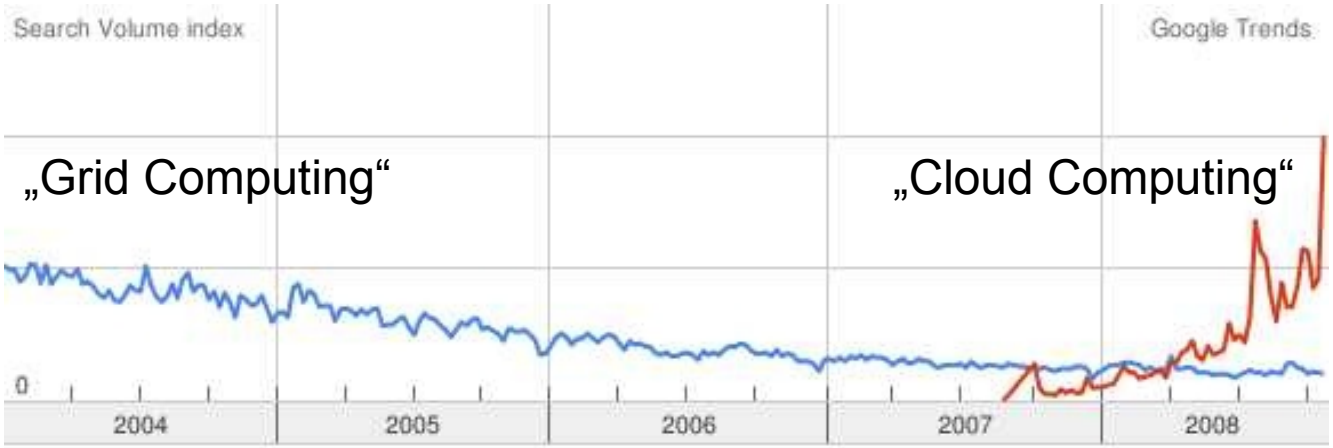
A Grid job example (5)

```
heiss@gridka25:~> glite-wms-job-output https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uy..
Connecting to the service https://wms-3-fzk.gridka.de:7443/glite_wms_wmproxy_server

================================================================================

                        JOB GET OUTPUT OUTCOME

Output sandbox files for the job:
https://lb-2-fzk.gridka.de:9000/P_NEqUnjyvz53uyIXDXTLw
have been successfully retrieved and stored in the directory:
/tmp/jobOutput/heiss_P_NEqUnjyvz53uyIXDXTLw

================================================================================



heiss@gridka25:~> cat /tmp/jobOutput/heiss_P_NEqUnjyvz53uyIXDXTLw/hello.out
nuni07
```

Cloud Computing

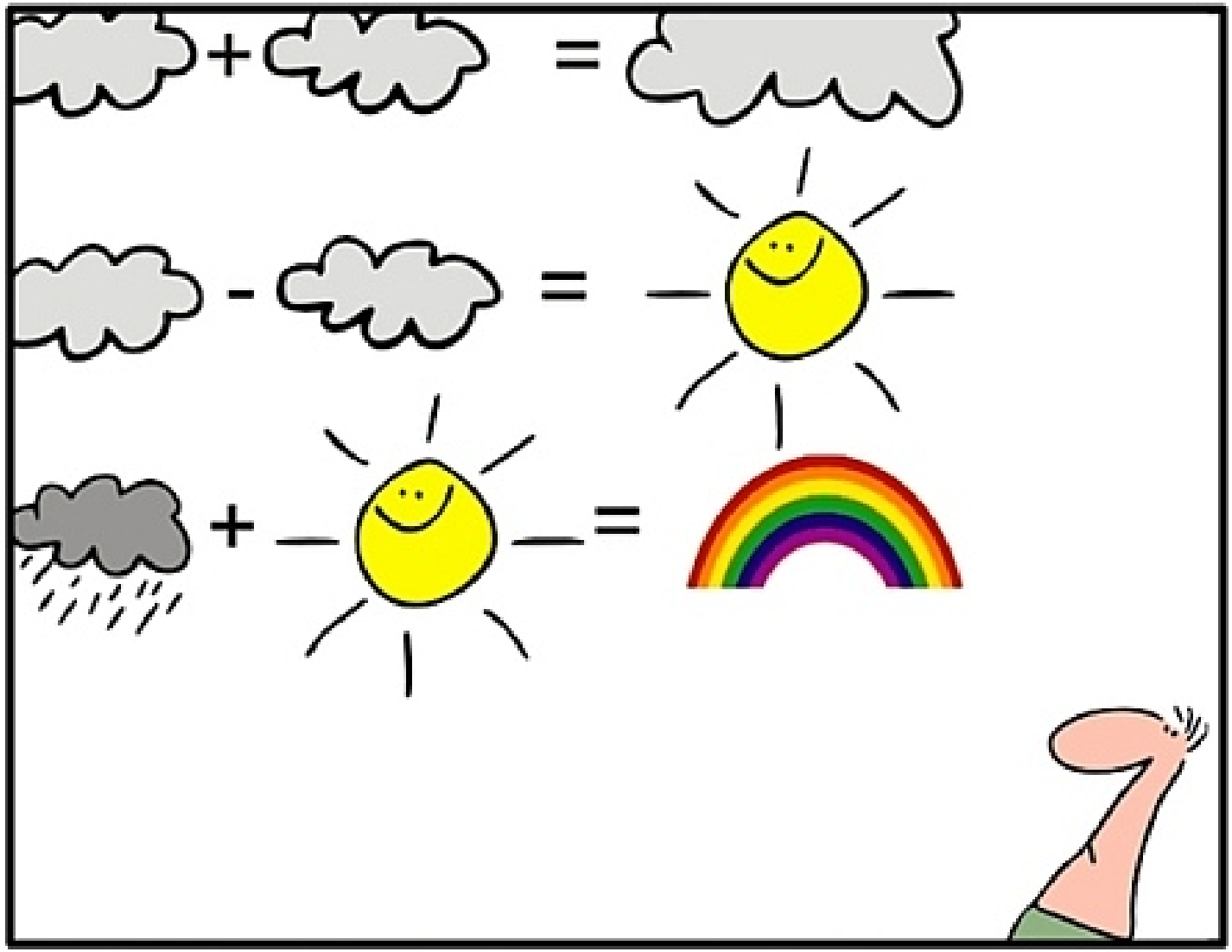
- The ”Grid” hype is over.
  - Grid is an established technique.
  - Not much used in industry but established in science
    - Main reasons: data security, Grids built for special user communities, usability

- New Hype: ”Cloud Computing”



Quelle: Google Trends

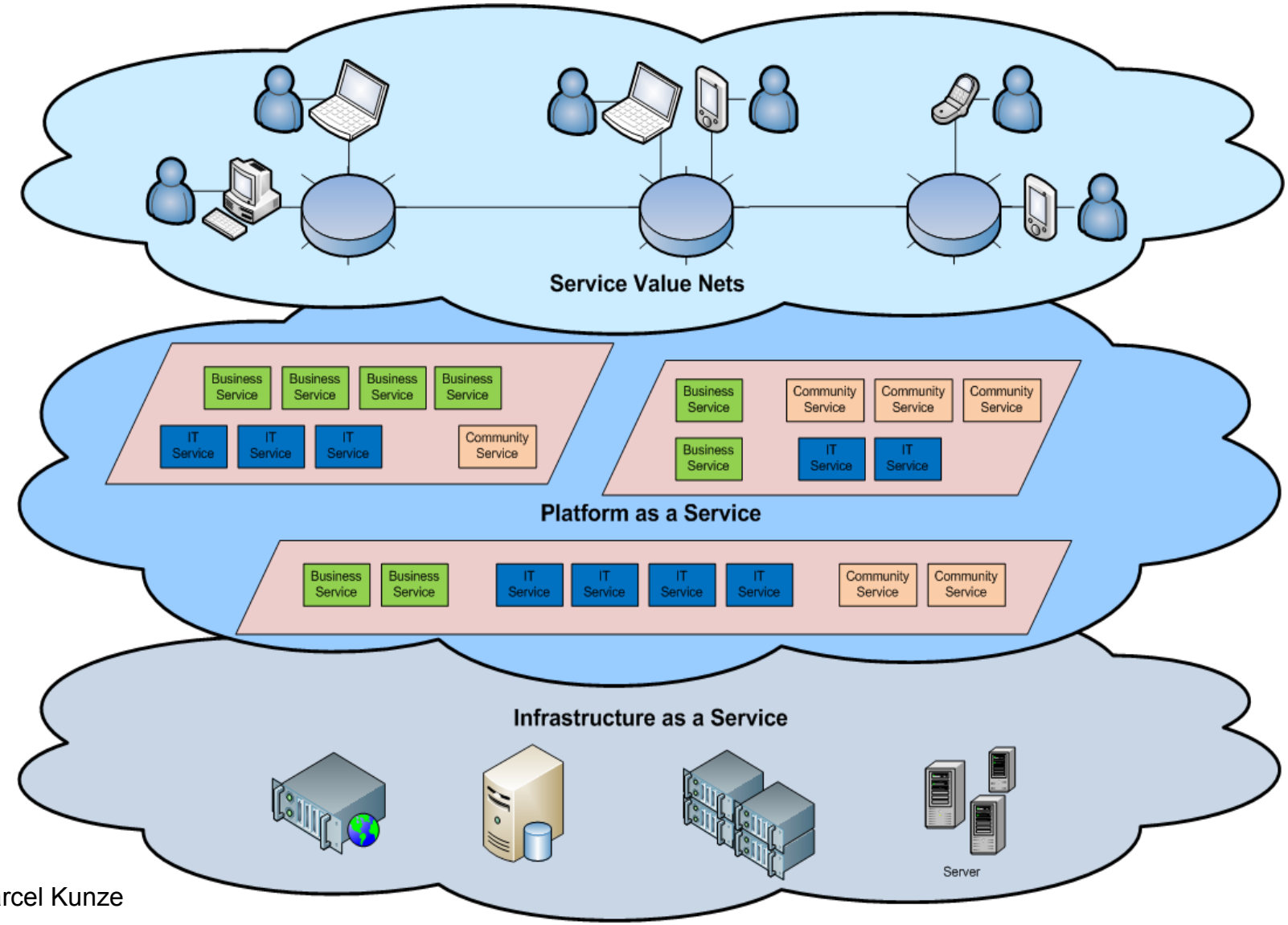
Cloud Computing



Cloud Computing

- Wikipedia: ”Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the Internet. ”
  - The concept generally incorporates combinations of the following:
    - infrastructure as a service (IaaS)
    - platform as a service (PaaS)
    - software as a service (SaaS)
- Amazon, Google, ”OpenCirrus” Cloud testbed
- Has its roots more in industry than in science
  - Problems to set global (open) standards?
  - Probably: creation of several stable and flexible (but incompatible) solutions → survival of the fittest?

Cloud Computing



Courtesy Dr. Marcel Kunze

Cloud Computing

- Cloud computing pursuits the same aims and visions as Grid computing

echo

“Grid computing is coordinated resource sharing and problem solving in dynamic, multi-institutional virtual organizations (I.Foster)“

| sed s/Grid/Cloud/g

Cloud Computing

- Ease of use: within minutes you can learn to use cloud resources
- Very flexible

- So far, automation is not on the same level than in Grids (global batch submission system)
- Large data volumes
- Service Levels? (Services distributed in the cloud / many cloud providers)

Karlsruhe Institute of Technology

Questions?